

OPPTRYKK AV
STATISTISK SENTRALBYRÅS HÅNDBØKER 44

INNFØRING I PRAKTISK STATISTIKK

STATISTISK SENTRALBYRÅ
OSLO 1985

ISBN 82-537-1118-2
ISSN 0085-4336

OPPTRYKK AV
STATISTISK SENTRALBYRÅS HÅNDBØKER 44

INNFORING I PRAKTISK STATISTIKK

STATISTISK SENTRALBYRÅ
OSLO 1985

ISBN 82-537-1118-2
ISSN 0085-4336

FORORD

Statistisk Sentralbyrå har i mange år drevet organisert opplæring av medarbeidere uten akademisk utdanning.

I faget praktisk statistikk har savnet av høvelige lærebøker vært følbart. For å rå bot på dette, har Byrået bedt byråsjef Tormod Andreassen om å gi et sammendrag av det stoffet han som lærer har gjennomgått på Byråskolens kurs III.

"Innføring i praktisk statistikk" er i første rekke ment å være et hjelpemiddel på Byråskolen og i Byrået ellers. Også andre som arbeider med eller har behov for å sette seg inn i praktisk-statistiske problemer, vil kunne ha nytte av innholdet.

Statistisk Sentralbyrå, Oslo, 12. mars 1980

Petter Jakob Bjerve

Gisle Skancke

	Side
I. EN DEL PRAKTISK-STATISTISKE GRUNNBEGREPER	7
1. Innledning	7
A. Hva er statistikk?	7
B. Hvem bruker statistikk?	7
C. Hvorfor brukes statistikk?	8
D. Hvem produserer statistikk?	9
2. Krav til statistikken	10
A. Innledning med definisjoner	10
B. Krav til omfang, ensartethet og sammenliknbarhet. Standarder	11
C. Engangsundersøkelser kontra løpende undersøkelser	12
D. Hyppighet av undersøkelser, aktualitet	13
E. Utvalgs- og totalundersøkelser	13
F. Økonomiske overveielser, kvalitet og tilgjengelighet	14
G. Internasjonalt statistisk samarbeide	15
3. Statistiske standarder	16
A. Behovet for standardisering	16
B. Standard for næringsgruppering	18
C. Vareklassifisering; Brussel-nomenklaturen og FN's standard	20
D. Standard for utdanningsgruppering	22
E. Standard for handelsområder	23
F. Standard for gruppering av sykdommer - skader - dødsårsaker	23
G. Andre statistiske standarder	24
4. Om sentrale statistiske registre	25
A. Hva er et register?	25
B. Hva brukes registre til?	26
C. Prinsipper for oppbygging og ajourhold av registre	27
D. De sentrale registre i Byrået	27
1. Det sentrale personregister	27
2. Bedrifts- og foretaksregisteret	28
3. Andre registre	30
5. Arkivstatistiske systemer	31
A. Arkivstatistikkens grunnide	31
B. Behovet for arkivstatistikk	31
C. Arkivsystemer - Databanker	32
D. Utnytting av dataarkiver	32
II. DE ENKELTE LEDDENE I EN STATISTISK UNDERSØKELSE	34
1. Spesifisering av formålet med undersøkelsen	34
A. Bakgrunnsstudier - hvordan skaffe observasjonsmaterialet?	34
B. Operasjonell utforming av formålet, systemering	37
1. Hva er systemering?	37
2. Hvordan systematisere datainnsamlingen?	40
2. Planlegging av gjennomføringen av undersøkelsen	40
A. Arbeidsplan, opplegg av undersøkelsen	40
B. Tidsplan	41
C. Kostnadsplan	42
3. Omfang, enheter og definisjoner	43
A. Identifisering av masse og enheter	43
B. Spesifikasjon av problemrelevante egenskaper	43
4. Oppgaveinnhenting	44
A. Metoder for innhenting, valg av tellemåte	44
1. Postundersøkelser	46
2. Telefon-, telegrafundersøkelser	47
3. Intervjuundersøkelser	48
4. Registerundersøkelser	48
B. Blankettarbeidet - skjemaer og spørsmålsformulering	49
5. Bearbeiding av observasjonsmaterialet	54
A. Revisjon, kontroll og oppretting	54
1. Manuell bearbeiding kontra EDB-bearbeiding	59
2. Manuell bearbeiding	60
3. Koding og dataoverføring	60
4. Maskinell revisjon med oppretting	62
B. Bearbeidings- og resultatkontroll	66
C. Standardisering	67

6. Elementær analyse	67
A. Gruppering og opptelling	68
1. Tabeller	68
2. Grafisk framstilling	71
B. Utregning av statistiske mål, gjennomsnitts- og spredningstall	83
C. Indekser	83
1. Konsumprisindeksen	83
7. Publisering og lagring av data	88
A. Publisering	88
1. Publiseringsmåter og kommentarer	88
2. Trykkeriteknikk	91
3. Spesialservice	92
B. Lagring	93
C. Dokumentasjon av data	94
1. Dokumentasjon beregnet på interne formål	94
2. Dokumentasjon som dekker brukerne av statistikken	97
Utkomne publikasjoner	
Utkommet i serien Statistisk Sentralbyrås Håndsbøker (SSH)	98

I. EN DEL PRAKTISK - STATISTISKE GRUNNBEGREPER

1. INNLEDNING

A. Hva er statistikk?

I faget statistikk tar en for seg prinsipper for innsamling, bearbeiding og presentasjon av tall og andre data. Faget omfatter også regler for bruk av tall og hvordan en kan ta beslutninger på grunnlag av tallene. Statistikken er et fagområde på samme måte som "matematikk", "sosiologi", "journalistikk" osv. Statistikk er altså ikke bare meningsfylte tall, men også metoder, prinsipper, teorier osv., dvs. statistikk inneholder alle de elementer som skal til for at vi kan kalle det et fagområde. Begrepet statistikk ble visstnok første gang brukt for 200 år siden av Gottfred von Achenwald, og han mente statistikk var det samme som kunnskap om samfunnet.

Det er av og til ønskelig å dele faget statistikk i to deler: 1) Talloppgaver over visse hendelser og 2) Det teoretiske og praktiske grunnlaget for å arbeide med talloppgaver. Del 2 deles videre ofte opp i to deler, nemlig teoretisk og praktisk statistikk. Teoretisk statistikk betegnes av og til som matematisk statistikk og praktisk statistikk som deskriptiv statistikk. Noe tydelig skille er det imidlertid ikke. Den teoretiske statistikken må sies å være en grein av den anvendte matematikken, og den er utviklet for å være til hjelp når vi skal tolke tallene, gi tallene en mening i forhold til det problemet vi skal belyse. Den anvendte eller praktiske statistikken går enda et skritt videre i retning av praktiske arbeidsproblemer.

I dagligtale, i aviser, i radio og andre massemedia er statistikk oftest bare en samling tall presentert i en tabell, i et diagram eller i en tekstlig kommentar. Bak tallene som presenteres ligger det mye arbeid av både praktisk og teoretisk art. Hensikten med å presentere visse statistiske tall, er på en komprimert måte å gi informasjon om enheter eller hendelser. En enhet kan her f.eks. være en person eller flere personer, like personer eller forskjellige personer. Seinere vil vi komme nærmere inn på hva vi i statistikken presist skal forstå med en enhet og andre statistiske begreper.

Vi ønsker altså å bringe fram informasjon ved hjelp av statistikk, statistisk informasjon. Med statistisk informasjon vil vi forstå systematisk tallmessig informasjon hvor de enkelte tall bygger på observasjoner av et sett av enheter, som vi mener det er hensiktsmessig å tenke oss samfunnet sammensatt av.

Når vi sier at statistikken er tallmessig informasjon for et sett av enheter, gir det uttrykk for at vi er interessert i informasjon om den enkelte enhet (person, bedrift etc.) bare i den hensikt å skaffe informasjon om hele grupper (f.eks. pensjonister, industrien osv.). Samfunnet må også få en vid tolkning her, det kan stå for verdenssamfunnet, et trossamfunn, en kommune osv.

I noen grad det teoretiske, men i første rekke det praktiske grunnlaget for å arbeide med talloppgaver (enkelttall), er det vi skal gå nærmere inn på i det følgende.

B. Hvem bruker statistikk?

Statistiske oppgaver brukes av mange forskjellige brukergrupper for en mengde ulike formål. Behovet for statistiske talloppgaver, eller m.a.o. systematisk tallmessig informasjon om visse enheter, vil variere med formålet. Brukerne kan hensiktsmessig inndeles i følgende hovedgrupper:

1. Storting og regjering
2. Sentraladministrasjonen
3. De kommunale og fylkeskommunale administrasjoner
4. Utredningskomitéer
5. Næringsliv og næringsorganisasjoner
6. Politiske og ideologiske organisasjoner
7. Forskningsinstitutter og forskere
8. Undervisning
9. Massemedia: kringkasting og presse
10. Interesserte enkeltpersoner
11. Internasjonale organisasjoner

Hensikten med å dele brukerne inn i slike hovedgrupper er at statistikkprodusenten da har mulighet til å gjøre statistikken mer brukerorientert ved at han kan henvende seg til brukergruppene og så undersøke deres behov for statistikk. Markedsføringen av statistikk har to sider: Selve produkt- - utformingen (hvordan bør publikasjonene se ut?) [layout] etc., og spredningen av statistikkproduktene (hvem bør ha de forskjellige statistikker?) [publikasjoner]. Valg av medium (publikasjon, maskinlister, magnetbånd, dataskjerm, mikrofilm) vil avhenge av statistikkproduktet og bruker. Statistikkproduktet (mediet) bør utformes og tilpasses brukerne fordi noen er interessert i statistikk for mange fagområder, andre i statistikk for bare ett fagområde. Noen er interessert i en grov oversikt, andre er interessert i detaljer. Noen vil ha aktuelle tall, andre er mer interessert i historien. På mange felter vil behovene fra de forskjellige hovedgruppene være sammenfallende. Hvorfor? Jo, fordi det kan godt tenkes at det er flere enn en gruppe som ønsker å få tallfestet akkurat de samme størrelser, enhetene. F.eks. ønsker både politikere, sentraladministrasjonen og næringslivet å få informasjon om personbilholdet, dvs. statistiske talloppgaver over antall biler, kjørevaner, teknisk tilstand osv.

Det er vanskelig, nærmest umulig, å tallfeste hvor mye statistikk hver av disse gruppene gjør bruk av, men som illustrasjon kan nevnes at Statistisk Årbok 1979 hadde et opplag på 35 000. Den offisielle statistikken består av langt flere publikasjoner enn Årboka, og i tillegg utgis en rekke private statistikker.

I Norge er utarbeidingen av den offisielle statistikken i høy grad sentralisert til Statistisk Sentralbyrå. Ifølge rundskriv fra Finansdepartementet (1972) er departementer og underliggende organer pliktige til å legge fram for Byrået skjemaer som planlegges sendt ut for innhenting av statistiske opplysninger.

Byrået får kjennskap til hvordan de ulike brukergruppene ønsker statistikken gjennom mange former for kontakt: Rådgivende utvalg, personlig kontakt, offentlige dokumenter, skriftlige henvendelser og bestillinger, seminarer, foredrag osv., og spesielle undersøkelser¹⁾.

C. Hvorfor brukes statistikk?

Statistikk brukes for å få informasjon. Brukerne har nemlig behov for informasjon fordi de skal

- 1) Løse aktuelle problemer og
- 2) Være i beredskap for å løse beslutningssituasjoner som oppstår.

Aktuelle problemer kan bli gjenstand for

- a) Forskning og utredning,
- b) Prøving og kontroll eller
- c) Umiddelbar løsning ved hjelp av statistisk informasjon.

Enkelte brukere må være i beredskap for å løse eventuelle situasjoner som kan oppstå eller som vi vet vil oppstå, f.eks. må arbeidskraftsmyndighetene være i beredskap når det gjelder utviklingen i arbeidsløshet. Slike situasjoner krever beslutninger, og statistisk informasjon er en viktig del av beslutningsunderlaget.

Et moderne samfunn trenger statistikk, av flere grunner. Statistikk er nødvendig for at vi skal få oversikt over situasjonen og utviklingen i økonomiske, sosiale, kulturelle og miljømessige forhold. Statistikk brukes når vi skal forklare sammenhenger mellom ulike sider ved utviklingen. Når noen skal løse et problem, enten det er en vitenskapsmann eller hvilken som helst annen intelligent problemløser, deler han bevisst (eller ubevisst) problemløsningen opp i fire deler:

1. Observasjon. Han observerer og samler fakta som er av betydning for problemet.

2. Hypotesen. Han ønsker å forklare sitt problem, dvs. angi årsaksforholdet i problemet, finne fram til de faktorer som forklarer og beskriver problemet. Sine forklaringer, som er basert på observasjoner (jfr. 1), stiller han opp i en hypotese.

1) I tiden 1. september - 30. november 1967 foretok alle kontorer i Byrået en registrering av den innenlandske etterspørsel etter statistiske data, og i hele 1967 foretok Byrået en registrering av avisenes bruk av de pressemeldinger som sendes ut i serien Aktuell statistikk. Resultatet av disse undersøkelsene er publisert i et arbeidsnotat, IO 68/17: Om bruken av SSB's statistikk og Paal Bog (Oslo, 30. juli 1968).

I et Arbeidsnotat, IO 68/8, er to statistikkbrukeres databehov presentert, nemlig "Departementenes databehov" og "Det private næringslivs og næringsorganisasjonenes databehov" (Statistikkbrukernes syn på SSB's statistikk, Oslo, 10. juni 1968).

3. Prediksjoner. Hvis hans hypotese er riktig, vil han kunne forutsi hvordan og når problemet/hendelsen oppstår. Prediksjonen går ut på å forutsi hva som vil skje dersom årsaksfaktorene han har funnet fram til, endres, dvs. får en annen verdi.

4. Verifisering. En testing av prediksjonen vil enten bekrefte eller avkrefte hypotesen. Prediksjonen testes ved å samle inn nye observasjoner av de faktorer som forklarer og beskriver problemet. Testen kan enten føre til forkastelse av hypotesen eller en verifisering. Blir hypotesen verifisert, er det ikke dermed sagt at årsaksforholdet er bevist, men det er større grunn enn før testingen til å anta at hypotesen er rett.

D. Hvem produserer statistikk?

Statistikk er altså tallmessig informasjon om visse hendelser. Denne tallmessige informasjonen kan gis av forskjellige produsenter: Offentlige institusjoner, enkeltpersoner, bedrifter, organisasjoner osv.

Her i landet er det Statistisk Sentralbyrå som har til oppgave å lage og legge til rette for brukerne det meste av den statistikken som har allmenn interesse og som går under navnet offisiell statistikk. Statistikk som offentlige organ, private foretak, organisasjoner eller andre trenger og utarbeider til egne formål, slik at de har liten eller ingen interesse for andre, faller utenfor Byråets virksomhet og betegnelsen offisiell statistikk.

En skiller ofte mellom primærstatistikk og sekundærstatistikk. Primærstatistikk er tallmessig informasjon utarbeidd for et bestemt statistisk formål, mens sekundærstatistikk er tallmessig informasjon som belyser et problem, en sak, uten at disse tallene er samlet inn direkte for å belyse akkurat dette statistiske problemet. Den primære hensikten med tallene har ikke vært å gi informasjon om vårt problem, men har som regel vært et ledd i en administrativ saks behandling. Primærstatistikk som bearbeides på forskjellige måter, f.eks. ved at man kombinerer tallopgaver fra forskjellige kilder, slik en f.eks. gjør i nasjonalregnskapet, fortjener da merkelappen sekundærstatistikk.

F.eks. er konsumprisindeksen en primærstatistikk; pristallene som samles inn blir samlet nettopp for å tallfeste konsumprisindeksen. Derimot er en indeks for reallønnen en sekundærstatistikk, fordi der blir en indeks for lønninger deflatert med konsumprisindeksen, og konsumprisindeksen (grunnlaget for utarbeidelsen av denne) blir jo ikke beregnet for at en skal komme fram til reallønnen. Konsumprisindeksen er bare det deflateringsmål som allerede utarbeides og som en har funnet best egnet.

Skattestatistikken derimot er sekundærstatistikk. Statistikken er utarbeidd på grunnlag av likningsoppgaver (selvangivelsen), og opplysningene skattyterne gir på selvangivelsen er tilpasset likningsvesenets behov. Likningsvesenet er neppe så interessert i å skille mellom de ulike fradragspostene som skattestatistikken er. Så lenge en utgift blir godtatt som fradragspost, er det av mindre betydning for likningsvesenet om det er spesifisert eller slått sammen med andre fradragsposter.

Statistikk eller tallmessig statistikk er representasjon av informasjon, og den som produserer slik informasjon er da enten pålagt det i henhold til lov (f.eks. Statistisk Sentralbyrå) eller de ønsker å informere om seg selv, om andre eller om spesielle hendelser.

Statistikk blir produsert enten som løpende periodiske tellinger eller i form av engangstelling (eventuelt flere ganger, men ikke regelmessig og periodisk). I publikasjonen Veiviser i offisiell statistikk, utgitt av Statistisk Sentralbyrå, gis en oversikt over hva som produseres av offisiell løpende periodisk statistikk i Norge i dag. I oversikten er det for den enkelte statistikk også angitt hvem som utarbeider den, hva den omfatter og periodelengden, dvs. hvor ofte statistikken utarbeides. Periodisk statistikk utarbeides ikke bare av offisielle organer, men også av bedrifter, organisasjoner o.l. av helt uoffisiell karakter. Slik ikke-offisiell statistikk utarbeides oftest til internt bruk, og denne statistikken offentliggjøres da heller sjelden. Ikke-offisiell statistikk er all slags tallmessig informasjon som ikke er tilgjengelig for alle; f.eks. en bedrifts interne statistikk (sykdomsfravær, salgsstatistikk osv.).

Statistikk produseres for å tilfredsstille et informasjonsbehov. Dette behovet kan sammenliknes med andre behov vi ønsker tilfredsstilt. Vi i denne sammenheng er enten en bedrift, en husholdning, staten, en kommune osv., og det er klart at vi ikke vil ha denne informasjonen uansett hva den koster. Det koster noe å frambringe informasjon, og disse utgiftene må dekkes. Produksjon av statistikk må konkurrere med annen tjenesteproduksjon når det gjelder bruk av ressurser og når det gjelder etter-spørrel. I prinsippet må dermed nytten av den informasjon vi får ved å produsere en bestemt statistikk være minst like stor som det koster å produsere denne statistikken. I praksis kan det være meget vanskelig å avgjøre om dette virkelig er tilfelle.

2. KRAV TIL STATISTIKKEN

A. Innledning med definisjoner

Produktene fra de forskjellige statistikkområdene skal tilfredsstille brukernes behov for informasjon om så vel strukturen som utviklingen innenfor forskjellige områder. Strukturelle data, tall, vil være verdier på kjennemerker ved de enheter som karakteriserer, beskriver, det statistikkområde eller problem vi er interessert i på et bestemt tidspunkt eller i en bestemt periode. F.eks. kan forhold omkring "Befolkningen i Norge" beskrives ved hjelp av en rekke kjennemerker for de enheter, dvs. personer, som utgjør befolkningen i Norge. Slike kjennemerker er kjønn, alder, ekteskapelig status osv. Fordelingen av personer etter verdier på disse kjennemerker (f.eks. antall menn i 30-årene som er gift, antall gifte kvinner i alt osv.) vil si noe om befolkningsstrukturen i Norge i et bestemt år. Utviklingen vil vi få når vi sammenlikner fordelingen for de samme kjennemerker fra et år til et annet (f.eks. tall for gjennomsnittsalderen ved inngåelse av ekteskap i 1976 kan sammenliknes med tilsvarende tall for 10 år tidligere).

Byrået er en statistikkprodusent, og konsumentene eller brukerne er alle de som bruker Byråets statistikk. Byrået bruker også selv i sin analysevirksomhet mye av denne statistikken.

Det å produsere statistikk koster både direkte og indirekte penger:

- 1) Direkte i form av lønninger og andre utgifter hos statistikkprodusenten,
- 2) Indirekte av oppgavegivere og andre som må besvare skjemaer og forespørsler fra produsenten.

Før man setter i gang med å produsere statistikk må man derfor nøye vurdere: a) Verdien av statistikken, dvs. hvilken avkastning produksjonen gir, b) Kostnadene ved å produsere denne statistikken.

Før vi går videre vil vi omtale et par begreper som vi allerede har nyttet noen ganger: En masse er en samling enheter som har visse felles kjennemerker. Massen, eller populasjonen, som den også av og til kalles, er alle de enhetene som vi vil dekke ved en statistisk undersøkelse. Massen vil derfor kunne variere fra undersøkelse til undersøkelse. Vi avgrensar eller samler enhetene i en masse fordi de har en rekke egenskaper felles. I den seinere tid har også begrepet Undersøkellesbestand¹⁾ blitt tatt bruk istedenfor begrepet masse. Vi vil bruke begge begrepene.

Massen kan i tillegg til en samling fysiske eller juridiske enheter også være en samling av hendelser, f.eks. tyverier, trafikkulykker osv. Ofte kan det derfor være hensiktsmessig å dele inn statistikken etter den type av observasjoner vi står overfor: 1. Eksperimentellstatistikk (f.eks. kast med terning, kvalitetsprøving av lyspærer, osv.). Vi har her kontroll over eksperimentet, 2. Massemålinger av gitte populasjoner (f.eks. rekruttenes høyde, industriens produksjon, osv.). "Noen" har gjort eksperimentet, vi registrerer resultatet, 3. Hendelsesobservasjoner (f.eks. tyverier, osv.).

Egenskapene, eller kjennemerkene, er det vi karakteriserer eller beskriver enhetene ved hjelp av. Enhetene på sin side karakteriserer massen (undersøkellesbestanden), og i statistisk sammenheng er det først og fremst hele massen og deler av massen vi er interessert i å få beskrevet, ikke hver enkelt enhet. For hver enhet i massen kan vi måle egenskaper vi er interessert i (f.eks. bedriftens sysselsetting, husholdningens forbruk av mat, prisen på ost, osv.). Egenskapene ved enhetene, kjennemerkene, kalles også kjennetegn eller variable. Verdien på et kjennemerke, variabelverdien, kalles oftest for observasjonen eller koden, av og til også for kjennemerkeverdien.

En masse består altså av enheter som har visse felles kjennemerker og visse ulike kjennemerker. Problemet vi vil løse vil da ofte være at vi skal finne ut hvorfor eller hvordan de forskjellige kjennemerkene finnes igjen hos de enkelte enhetene. (F.eks. vil alle industribedrifter i Norge være en interessant masse som kan beskrives ved hjelp av kjennemerket sysselsetting, og antall sysselsatte, ca. 400 000, vil da være kjennemerkeverdien.) En enhet kan vi også karakterisere ved å plassere den i en gruppe, klassifisere enheten. En klasse vil da som regel være angitt med to verdier på en eller flere kjennemerker. (F.eks. kan en tenke seg husholdningen delt opp i to grupper: Med og uten barn. Verdien på kjennemerket "antall barn" vil da plassere husholdningen i den ene eller andre gruppen.)

1) Se "Noen betraktninger om dataterminologi". Arbeidsnotat av Jan M. Hoem. IO 75/26.

Endelig vil vi se på forskjellen mellom klassifikasjonsenhet og rapporteringsenhet. En statistisk undersøkelse utføres for å besvare visse problemstillinger. Problemstillingene kan være knyttet til forskjellige typer av enheter, klassifikasjonsenheter. (Enheter blir gruppert på grunnlag av klassifikasjonsprinsipper, f.eks. klassifisert som bedrift eller som foretak.) Den som besvarer spørsmålene (rapporteringsenheten), kan godt besvare spørsmål for flere forskjellige typer av enheter (f.eks. kan foretaket besvare spørsmål om bedriftene i foretaket og om hele foretaket; eller hovedpersonen i en husholdning kan besvare spørsmål om hele husholdningen og dessuten om hovedpersonen alene). - Rapporteringsenheten (svareren) kan altså godt besvare flere spørsmål (gi verdier på kjennemerker) for flere enn én klassifikasjonsenhet (enhetene det blir undersøkt noe om).

Klassifikasjonsenhet kalles av og til telleenhet, og spesielt når det ikke er forskjell på klassifikasjonsenheten og rapporteringsenheten, brukes bemerkelsen telleenhet.

Seinere vil vi ved hjelp av flere eksempler komme nærmere inn på masse og enhet.

B. Krav til omfang, ensartethet og sammenliknbarhet. Standarder

Når det gjelder en bestemt statistikkens omfang, må produsenten av denne statistikken ta stilling til følgende spørsmål:

1. Hvilken bredde skal statistikken ha? Hva skal dekkes av akkurat denne statistikken, dvs. hvilke områder og hvilke enheter skal dekkes? M.a.o. en avgrensing av masse og enheter
2. Hvor detaljert skal vi lage statistikken? Hvilke kjennemerker ved enhetene er det vi skal undersøke og lage statistikk for?
3. Hvor ofte skal vi spørre? Skal vi spørre en gang for alle, eller skal vi spørre med jamne mellomrom?
4. Prioritere enkelte spørsmål framfor andre hvor kostnadene og/eller andre faktorer setter "et tak" (en grense) for hva vi kan ta med

La oss ta et eksempel: Anta at vi er interessert i industrien i Norge, og vi skal lage en statistisk undersøkelse av industrien. Vi må da bl.a. ta stilling til hva som skal regnes for industri-bedrifter (jfr. pkt. 1 ovenfor - er byggmestere, ferdighusprodusenter osv. industribedrifter?); er vi f.eks. interessert i bare antall ansatte i alt eller er vi også interessert i antall kvinner og menn (jfr. pkt. 2); vil vi vite hvordan det går med industrien fra måned til måned eller er det nok en gang i året (jfr. pkt. 3)? Svarene på spørsmålene 2 og 3 vil si om vi er mest interessert i endringer over tid (og hvor lang tid) eller om vi er mest opptatt av strukturen på bestemte tidspunkter.

Statistikkproduksjonen stiller et hovedkrav til oppgavene, nemlig at oppgavene fra de ulike oppgavegiverne må være ensartet. Et homogent grunnmateriale er nødvendig for at aggregerte¹⁾ størrelser skal kunne gis en fornuftig tolkning. Husk det er mulig å aggregere detaljer, men det er ikke mulig å dele opp aggregerte størrelser (f.eks. kan vi spørre om antall ansatte menn og kvinner hver for seg og av dette også få fram antall ansatte i alt, men hvis vi spør bare om antall ansatte, kan vi ikke splitte dette opp - etter at vi har fått oppgavene - i menn og kvinner). Men detaljer koster penger.

Ved hjelp av statistikk skal vi belyse aggregerte størrelser eller sett av enheter som vi sa i innledningen. Det er da klart at ber vi om visse data fra disse enhetene, må dataene representere (gi tallmessig informasjon om) de samme kjennemerkene for alle enhetene og ikke fra forskjellige kjennemerker for noen av enhetene.

Forskjellige oppgaver kan vi få fordi spørsmålene er dårlig utformet eller det spørres om noe som ikke alle oppgavegiverne kan svare på. Er spørsmålene utformet slik at de gir rom for forskjellige tolkninger, er de dårlig utformet hvis det er bare en tolkning spørsmålsstillerne har hatt i tankene. Det er likevel ofte litt av et problem å få spørsmålene riktig formulert slik at oppgavegiverne gir mest mulig entydige svar.

1) En aggregert størrelse, et aggregat, vil si at vi har summert over kjennemerkeverdiene for ett eller flere kjennemerker for en gruppe enheter. Også andre samle størrelser enn en sum kalles for aggregater, f.eks. gjennomsnitt.

Skal vi sammenlikne oppgaver hentet inn på forskjellige tidspunkter, må vi være klar over om enhetene vi sammenlikner er definert på samme måte på de to tidspunkter eller om kjennmerkene vi sammenlikner verdier på, er definert likt. Når vi skal analysere utviklingen innenfor et bestemt område, sammenlikner vi tall fra forskjellige tidspunkter. Tallene må da gi informasjon om det interessante i denne sammenlikningen, og da må vi tallfeste de samme kjennmerkene for like enheter. Vi beskriver en enhet ved hjelp av kjennemerker. For å sikre at vi får opplysninger fra de samme enheter og om de samme kjennmerkene fra disse enhetene, er det at vi utarbeider standarder. Det er klart at vi også kan foreta sammenlikninger¹⁾ av data om forskjellige enheter, f.eks. sammenlikne industribedrifter med varehandelsbedrifter; bedrifters inntekter med personers inntekter, osv.

En standard er således noe som utarbeides for å sikre standardisering av de enheter, kjennemerker, ord og begrep som vi gjør bruk av i statistiske undersøkelser. Hovedhensikten med en standard for enheter og kjennemerker er at det ved alle grupperinger av disse enhetene og kjennmerkene så vidt mulig blir brukt samme klassifikasjonsprinsipper. På den måten kan opplysninger hentet fra ulike grupper av statistikken og fra ulike tidspunkter sammenliknes uten videre og gjensidig kontrolleres mot hverandre. Statistikkens informasjonsverdi vil da være større ved slike sammenlikningsmuligheter enn uten.

Helt sentralt i forbindelse med utarbeidelse av standarder står klassifikasjonsprinsippene. En standard lar seg vanskelig bygge opp på grunnlag av bare ett enkelt klassifikasjonsprinsipp. Hvilke prinsipper som nyttes i de enkelte tilfelle vil være både et valg mellom hensiktsmessighet og formål med de eller den statistikk som skal standardiseres. En vil forsøke ved hjelp av en standard å komme fram til interessante grupperinger av enhetene og kjennmerkene. Interessante i den forstand at de sier noe om 1) Det aktuelle problemet, eller 2) Gjør oss i stand til å løse beslutningssituasjoner som kan oppstå.

Byrået har utarbeidd en rekke slike standarder for en del statistikkområder. Vi skal seinere se litt nærmere på noen av disse.

C. Engangsundersøkelser kontra løpende undersøkelser

En statistisk undersøkelse kan enten være en løpende undersøkelse eller en engangsundersøkelse. Løpende undersøkelser gir oss løpende statistikk, dvs. statistikk som stadig gjentas. Det er et fast tidsrom, en periode, mellom hver gang undersøkelsen gjentas. Perioden kan være et år (årsstatistikk), en måned (månedssstatistikk), en uke osv. Blir periodelengden på flere år, f.eks. 10 år som er tilfelle for bl.a. folketellingene, kalles statistikken gjerne en engangsstatistikk selv om tilsvarende statistikker har vært utarbeidd tidligere. I motsetning til løpende statistikk må man ved engangsstatistikk bygge opp en personalstab som må læres opp, utstyres med hjelpemidler etc. bare for denne ene statistikken. Etter hvert som denne staben blir ferdig med produksjonen av statistikken, er også arbeidet slutt for dem. Ved løpende undersøkelser har man en fast stab.

Formålet med en engangsundersøkelse er å få klarlagt strukturen på et bestemt tidspunkt hos de enhetene man vil ha informasjon om. Strukturen belyser vi ved hjelp av kjennemerker som viser oss de karakteristiske egenskapene hos enhetene. Strukturelle kjennemerker er som regel kjennemerker hvor verdiene er stabile, dvs. endrer seg lite over tiden. En bruker begrepet struktur (og dermed strukturelle kjennemerker) for forhold hvor tidsaspektet ikke er sentralt.

Med løpende undersøkelser vil vi ønske å få informasjon om endringer i kjennmerkene (dvs. verdien av kjennmerkene) fra et bestemt tidspunkt til et annet. Det kan godt være de samme kjennemerker som vi ønsker å tallfeste ved engangsundersøkelsene som ved de løpende undersøkelsene, men ofte vil det være andre problemstillinger som skal løses ved hjelp av tidsseriestatistikk enn ved hjelp av tverrsnittsstatistikk. (Tverrsnittsstatistikk produseres som regel ved hjelp av engangsundersøkelser, mens tidsseriestatistikk produseres ved hjelp av løpende, periodiske undersøkelser.)

1) Sammenliknbarhet er utførlig behandlet av Hoem i ovenfor siterte Arbeidsnotat (ANO IO 75/26).

D. Hyppighet av undersøkelser, aktualitet

Kravet til aktualitet vil ofte stride mot kravet til omfang. Et omfangsrikt materiale tar det lengre tid å innsamle, bearbeide og publisere enn et mindre materiale. Ofte må det foretas en avveining mellom det å få statistikken hurtig ut og at den skal være omfattende. Det er ikke alltid at dette valget må foretas. Av og til må vi ha informasjon om visse enheter før et bestemt tidspunkt, og da kan det hende at omfanget nærmest "bestemmer seg selv".

Det er visse problemer som kan behandles bare ved hjelp av hyppige undersøkelser. Tilsvarende for beslutningssituasjoner (f.eks. konjunkturproblemer og konjunkturpolitikk).

Enheter med kjennemerker som kan få sine verdier hurtig endret og hvor endringene vil kunne ha stor betydning for samfunnet, for organisasjoner osv. vil kreve hyppige undersøkelser. F.eks. er arbeidsføre personer enheter som det er viktig for landets økonomi å følge. Disse personene kan kjennetegnes ved om de er sysselsatte eller ikke. Økes antallet som er arbeidsledige, vil visse tiltak for å begrense arbeidsløsheten kunne bli satt i verk. Endringer i arbeidsløsheten kan skje fort, og det er derfor viktig å undersøke sysselsettingssituasjonen ofte.

Som konklusjon kan vi si at det er selve problemet som vil bestemme hvor ønskelig det er med hyppige undersøkelser. Behovet for aktuelle tall, som hyppige undersøkelser oftest vil føre til, vil måtte avveies mot andre hensyn, først og fremst omfanget, men ofte også slike ting som kvalitet og nøyaktighet. Jo lengre tid vi bruker til å heve kvaliteten på et datamateriale (bearbeide det grundig, kontakte oppgavegiverne hvis vi finner feil osv.), jo mindre aktuelt vil det bli.

E. Utvalgs- og totalundersøkelser

I enhver statistisk undersøkelse har en å gjøre med en (telle)enhet. Med dette mener en den enkelte enhet det søkes opplysning om. Enhetene kan være den enkelte bedrift, husholdning, bolig, gårdsbruk eller person. Skal en undersøkelse være pålitelig og vellykket, må enheten være klart avgrenset og definert.

Det er imidlertid ikke opplysningene om den enkelte enhet isolert som interesserer statistikkprodusenten eller statistikkbrukeren, det er opplysninger om en masse av enheter. En slik statistisk masse kan f.eks. være alle bedrifter i landet, alle husholdninger, alle boliger osv.

Når opplysninger samles inn fra samtlige enheter i massen, har vi med en totalundersøkelse - en fullstendig undersøkelse, eller fullstendig telling - å gjøre. Tar vi derimot for oss bare en del av enhetene, snakker vi om en utvalgsundersøkelse.

Fordelene ved å nytte utvalgsundersøkelser istedenfor fullstendige tellinger er blant annet at det er atskillig billigere. Ved gitt kostnadsramme kan en da ved en utvalgsundersøkelse legge mer arbeid i kontrollen av det enkelte skjema, i noen tilfeller f.eks. ved å hente inn manglende opplysninger - og sist, men ikke minst, resultatene kan komme raskt fordi en har å gjøre med et begrenset antall skjemaer.

Den viktigste årsaken til at vi velger å foreta en utvalgsundersøkelse er at kostnadene ved en totaltelling er for store. Av og til er det også helt unødvendig å undersøke hele massen fordi hele massen beskrives godt nok ut ifra vår problemstilling ved hjelp av bare et utvalgt antall enheter.

I Wallis and Roberts: "Hva er statistikk?"¹⁾gis det fem mulige grunner for å bruke utvalg istedenfor fullstendige tellinger.

- 1) Selve målingene av enhetene kan ødelegge enhetene som måles, derfor bør vi bare måle et utvalg
- 2) Hvis vi foretar en fullstendig telling, vil vi i alminnelighet få større nøyaktighet enn ved å foreta en utvalgstelling, men merutgiftene ved å foreta totaltelling i forhold til utvalgstelling vil av og til være større enn nytten av den bedre nøyaktigheten
- 3) Ved en utvalgsundersøkelse kan vi måle enhetene, dvs. angi verdier på kjennemerkene til enhetene, mer nøyaktig enn ved en totaltelling. Vi går da ut ifra gitte kostnader

1) Fakkeltbok F 106, Oslo, sidene 118-120.

- 4) Av og til er massen for stor, det er i virkeligheten uendelig antall enheter, slik at vi må ta utvalg.
- 5) Det hender at hele massen ikke er tilgjengelig, dvs. vi vil karakterisere en masse ved hjelp av enheter vi i dag ikke kan måle, og dermed vil vi måtte enten ta utvalg eller betrakte alle de tilgjengelige enhetene som et utvalg.

Vi kan for egen regning føye til et pkt. 6) Aktualitet. Det er raskere å bearbeide et lite datamateriale enn et stort.

Utvalgsundersøkelser, eller stikkprøveundersøkelser, har hatt og vil få stigende anvendelse. Små land som Norge vil imidlertid ikke med samme relative fordel kunne anvende en slik innsamlingsteknikk som de store befolkningsrike land kan. Det er alminnelig anerkjent at i de tilfelle der det har praktisk betydning, er det utvalgets totale størrelse og ikke utvalgsprosenten (den relative andel av den totale masse) som er avgjørende for nøyaktigheten. Et utvalg på 1 000 enheter av en masse på i alt 200 000 enheter, vil gi stort sett samme nøyaktighet som et utvalg på 1 000 enheter av i alt 20 000 enheter. Er det totale antall enheter (massen) stort, behøver utvalget derfor bare utgjøre en liten andel av den. For å oppnå pålitelige resultater må det da foreligge en utvalgsplan som bygger på statistisk-vitenskapelige metoder.

For utvalgsundersøkelser som skal foretas gjennom personlig intervju, må utvalget av intervjuobjekter trekkes fra geografisk begrensede områder. Dette av hensyn til de lokale medarbeidere som skal foreta intervjuet. Den utvalgsplan som nyttes ved Underavdelingen for intervjuundersøkelser i Statistisk Sentralbyrå er bygd opp i samsvar med disse metoder.

Ved undersøkelsene blir utvalgene av intervjuobjekter trukket i to trinn. Til dekningen på første trinn nyttes en inndeling av landet i om lag 400 utvalgsområder. Utvalgsområdene er stort sett kommuner, bortsett fra at kommuner med færre enn 3 000 innbyggere er slått sammen med andre kommuner.

Utvalgsområdene er først gruppert etter landsdel og region. Deretter er utvalgsområder (dvs. kommuner) innenfor samme landsdel og region som har et visst fellesskap i geografisk beliggenhet og næringsgrunnlag slått sammen i såkalt strata. Byer med mer enn 30 000 innbyggere er egne strata. På denne måten er landet delt opp i 102 strata. Fra hvert stratum (som altså er én eller flere kommuner) blir det på første trinn trukket ett utvalgsområde (altså én kommune).

På annet trinn blir det innen hvert uttrukne utvalgsområde trukket intervjuobjekter på grunnlag av et register over objektene (enhetene). For hvert av utvalgsområdene finnes et register over alle som til enhver tid er bosatt innenfor de geografiske grensene av utvalgsområdet.

Hver enkelt av de uttrukne enheter (oppgavegivere) representerer da en rekke andre enheter med omtrent de samme egenskaper som den datamaskinen tilfeldig har funnet fram til.

Utvalgsundersøkelser som ikke baseres på personlig intervju, men f.eks. på kontakt via posten, er ikke avhengig av at enhetene trekkes fra geografisk begrensede områder.

Det karakteristiske ved en vel planlagt og gjennomført utvalgsundersøkelse er at de resultater en kommer fram til har gyldighet for massen som helhet - såframt en vesentlig del av de uttrukne oppgavegiverne blir med i undersøkelsen. Noen anledning til å "bytte ut" uttrukne uvillige oppgavegivere med mer samarbeidsvillige har verken den som intervjuer eller den som planlegger undersøkelsen.

F. Økonomiske overveielser, kvalitet og tilgjengelighet

Som vi tidligere har påpekt, koster det å lage statistikk. Foruten at det koster å lage statistikken, dvs. produsere statistikken, koster det også noe for brukeren(e) å motta den informasjonen. Brukerens kostnader, oftest i form av arbeidstid han må bruke til å lese og forstå de statistiske oppgavene, vil i en viss grad være avhengig av hvor mye kostnader produsenten har lagt ned i produktet for å gjøre det tilgjengelig for bruker. Det er lettere å lese en figur med flere farger enn med to farger, men det koster noe mer å produsere en flerfarget figur. Det er ikke alltid at det koster noe særlig mer å lage en statistikk brukervennlig. Statistikkprodusenten må derfor også legge vekt på dette siste trinnet i informasjonsformidlingen.

Statistikkprodusenten må ikke begrense seg til å utgi et produkt som tilfredsstillende bare hans egne produsentkrav (innsamlingskrav, bearbeidingskrav osv.), men også brukers krav. I sin presentasjon av produktet må produsenten ta hensyn til hva det koster for brukeren å motta informasjonen. Det er mange måter å analysere og kommentere på, og derfor må brukers forståelse og tilegnelsestid for disse analyser og kommentarer vurderes av produsenten.

En statistikkens kvalitet kan bety forskjellige ting. Vi vil imidlertid her si at en statistikk har en høyere kvalitet enn en annen når informasjonsverdien er høyere. Det er dermed brukerens behov og verdsetting som angir kvaliteten.

Statistikken er tilgjengelig for brukerne gjennom publikasjoner og andre meldinger. Publikasjonene må derfor være utformet på en brukervennlig måte. Det bør være lett å finne fram i publikasjonene, og de forskjellige publikasjoner bør være koordinert slik at det er samme kjennemerkeverdi på samme kjennemerke over alt. (F.eks. gir Nasjonalregnskapet en koordinert framstilling av all økonomisk statistikk. Eksempler på andre publikasjoner hvor en har forsøkt å koordinere data fra mange kilder, er Sosialt utsyn 1974 og Miljøstatistikk 1976, begge utgitt i serien Statistiske analyser.)

G. Internasjonalt statistisk samarbeide

FN

FN er det høyeste organ også når det gjelder internasjonalt statistisk arbeid. Øverst har vi Statistical Commission. Denne kommisjonen er en av syv som ligger under FN's økonomiske og sosiale råd (ECOSOC). Den statistiske kommisjonen består av 18 medlemmer. De store landene som USA, England, Sovjet og Frankrike er faste medlemmer, de resterende skiftes mellom de mindre land. Ett av de skandinaviske land er som regel med. Kommisjonen møtes i 14 dager en gang hvert annet år for å trekke opp retningslinjer. Under the Statistical Commission hører the Statistical Office, New York (UNSO). Ved dette kontoret er det ansatt fast vel 150 personer fra forskjellige land. Det er ved dette kontor det statistiske arbeid foregår. Når et vedtak går igjennom i Statistical Commission og blir foreslått anvendt, er det vanlig at deltakerlandene forsøker å innordne seg slik at statistikken blir sammenliknbar (standardisering).

Under the Statistical Office sorterer alle de statistiske kontorer som hører til i de forskjellige Regionale økonomiske kommisjoner. Den regionale kommisjon for Europa (ECE) har kontor i Genève. Representanter fra USA, Canada og Israel møter også her, fordi disse landene ligger på samme utviklingsnivå som de europeiske landene. En gang årlig møtes alle sjefene fra hvert statistisk byrå for å diskutere saker av felles interesse (1-2 uker).

ECOSOC heter altså FN's råd for økonomiske og sosiale saker og har bl.a. statistikk-koordinering som oppgave. Formålet til dette rådet er bl.a.:

1. Fremme den statistiske utvikling i alle land
2. Være rådgiver for FN i statistiske spørsmål
3. Samordne fagorganenes statistikker
4. Lede den statistiske virksomhet ved FN's sekretariat

Direkte under FN finnes fem økonomiske kommisjoner som har ansvar for sitt geografiske område:

- ECE Europa (Genève)
- ECAFE Asia og Fjerne østen (Bangkok)
- ECLA Latin-Amerika (Santiago)
- ECA Afrika (Addis Abeba)
- ECWA Midtøsten (Beirut)

Europa er vel utviklet, det stilles derfor store krav til vår statistikk.

FN har mange datter- eller særorganisasjoner, bl.a.:

- FAO Food and Agriculture Organization i Roma
- IMF International Monetary Fund (Penge- og valutafondet) i Washington
- ILO International Labour Office (Det internasjonale arbeidskontor) i Genève
- ICAO International Civil Aviation Organization (Sivil luftfart)
- UNESCO FN's organisasjon for undervisning og kultur, som holder til i Paris
- UNIDO FN's organisasjon for industriell utvikling, som holder til i Wien
- WHO Verdens helseorganisasjon (World Health Organization)

Alle datterorganisasjonene har egne statistikk-kontorer.

Ofte kommer det forslag til endringer fra sekretariatet i f.eks. ECE. Arbeidsgruppen i ECE legger dette forslaget fram for sjefene når de møter i ECE. De sender det så videre til sekretariatet og derfra til kommisjonen i New York. Dette tar tid.

OECD

OEEC startet i 1948 som en følge av Marshallhjelpen, og ble seinere hetende OECD (Organization of Economic Co-operation and Development), som bl.a. kommer med årlig oversikt over den økonomiske utvikling i de forskjellige land. OECD består av de vesteuropeiske land,¹⁾ Hellas, Tyrkia, USA, Canada, Australia, New Zealand, Japan og Jugoslavia (assosiert). Sør-Korea har uttalt at landet vil søke om medlemskap i 1980.

Som en følge av energikrisen i 1973/74 er det opprettet et internasjonalt energisamarbeide IEA (International Energy Agency), som Norge er assosiert medlem av. IEA utarbeider energistatistikk og har en forholdsvis selvstendig stilling vis-à-vis sin moderorganisasjon OECD.

Nordisk statistisk samarbeid

Det arrangeres et sjefsmøte hvert 3. år hvor retningslinjene for samarbeidet diskuteres. De nedsetter utvalg for de forskjellige områder med medlemmer fra de forskjellige land (f.eks. Nordisk utvalg for næringsstatistikk, for prognosespørsmål osv.). Resultatet legges så fram for sjefene. Norden prøver å stå som en enhet utad.

Det foregår også en del uformell kontakt mellom de nordiske land.

Nordisk råd står bak oppgavene til FN's publikasjoner fra Norden samlet.

For å koordinere det statistiske arbeidet ble det i 1969 opprettet et permanent nordisk sekretariat for statistikk som har sete i København. Sekretariatet består nå av 3 saksbehandlere.

Viktigst for Norge er det nordiske samarbeid og FN. Det nordiske samarbeid består av de fem nordiske landene (Norge, Sverige, Danmark, Finland og Island), men Island deltar bare i mindre grad. De nordiske land prøver i stor utstrekning ved internasjonale møter å bli enige om sine internasjonale standpunkter, slik at man skal kunne sammenlikne statistikken for de nordiske land. Det nordiske statistiske samarbeidet har blitt noe mer komplisert etter at Danmark gikk inn i EF, som delvis gjør bruk av egne prinsipper og standarder som går på tvers av FN's anbefalinger på en rekke områder.

3. STATISTISKE STANDARDER

A. Behovet for standardisering

Opplysninger om enkelte enheter - hentet inn fra forskjellige kilder, for forskjellige formål og/eller på forskjellige tidspunkter - ønsker vi på forskjellige måter å sammenlikne med hverandre. Skal dette være mulig, må de statistiske enhetene defineres, og kjennemerkene bør ordnes i grupper på en systematisk måte. Ved siden av standardiserte gruppeinndelinger er det også nødvendig at de alternative verdiene som kjennemerkene kan anta, spesifiseres etter klart definerte regler og normer.

En statistisk standard er et instrument for å få til sammenliknbarhet over tiden, sammenliknbarhet mellom ulike statistikker og for å sikre en rasjonell innsamling, bearbeiding, publisering og anvending av statistisk informasjon.

I de neste to kapitlene skal vi gjennomgå de viktigste standarder som Byrået gjør bruk av. Hver standard vil bli gjennomgått ved at vi først tar for oss formålet med standarden, dvs. omtaler hva standarden skal definere og klassifisere (inndelegge). Deretter vil vi komme inn på klassifikasjonsprinsippene som er lagt til grunn for standarden før vi til slutt tar for oss noen av de viktigste definisjoner og klassifikasjoner som standarden inneholder. Framstillingen presenterer standarder som i første rekke sikter mot en enhetlig klassifisering av statistisk informasjon, og de kalles derfor statistiske klassifikasjonsstandarder. I norsk statistisk terminologi er det ikke skilt mellom nomenklatur og standard. Begge deler brukes som betegnelser på dokumenter som inneholder regler, inndelinger og anvisninger for klassifisering.

1) Belgia, Danmark, Finland, Frankrike, Irland, Island, Italia, Luxembourg, Nederland, Norge, Portugal, Spania, Storbritannia, Sveits, Sverige, Vest-Tyskland og Østerrike.

En statistisk klassifikasjonsstandard skal gjennom klassifisering beskrive enheter som individer, bedrifter, varer, trafikkulykker, priser osv. på grunnlag av enhetenes egenskaper (kjennermerker) og relasjoner (relasjoner mellom enhetene innbyrdes og relasjoner overfor "andre" enheter). Det alminnelige klassifikasjonsprinsippet er at en enhet klassifiseres i den klassen (gruppen) i standarden som best tilfredsstillende enhetens egenskaper (kjennermerker) innenfor standardens område. F.eks. vil en næringsklassifisering av en bedrift si at vi klassifiserer bedrift i den (nærings-)gruppen som passer best ifølge næringsstandardens klassifikasjonsprinsipper.

Når noen gjør bruk av en klassifikasjonsstandard, er det som regel et uttrykk for at man ønsker å begrense antall varianter av inndelinger og grupperinger innenfor et statistikkområde.

Formålet med en standard - samordning og øket sammenliknbarhet - stiller to på mange måter motstridende krav til standarden: På den ene siden bør standarden bestå gjennom lengre tid, mens den på den andre siden bør kunne tilpasses endrede ytre forhold. Utarbeiding av standarder setter derfor spesielle krav til framtidsrettethet hos forfatterne.

I Norge er den offisielle statistikken sentralisert til Statistisk Sentralbyrå, og som sådan er Byrået ansvarlig for samordning av den offisielle statistikken. Byrået skal også medvirke i det internasjonale statistiske arbeidet og passe på at norsk offisiell statistikk er tilpasset rekommendasjoner vedtatt av internasjonale statistiske organisasjoner. Byrået utarbeider de norske statistiske standardene, og det er ment at standardene skal brukes i all offisiell statistikk, men gjerne også i annen statistikkproduksjon og -anvendelse.

I det følgende er det gitt en gruppering av de mest vanlige statistiske standardene:

1. Generelle klassifikasjonstermer

- a) FN's nasjonalregnskapssystem (SNA - A System of National Accounts)
- b) Det norske nasjonalregnskapssystemet (bygger på SNA)
- c) FN's system for sosiale indikatorer (SSDS - A System of Social and Demographic Statistics; ennå ikke fastlagt som standard)

2. Klassifisering av produkter

- a) FN's varenomenklatur (SITC, Rev. 2 - Standard International Trade Classification - som tas i bruk i 1976)
- b) FN's klassifisering av varer etter næringen de er produsert i (ICGS - International Standard Classification of All Goods and Services, foreløpig ikke fastlagt som standard)
- c) Tolltariffen eller Brussel-nomenklaturen (BTN - Brussels Tariffs Nomenclature), fra høsten 1976 også betegnet med bokstavene CCCN
- d) Felles nordisk statistisk varenomenklatur (Gemensam nordisk näringsstatistisk varuförteckning, den siste er fra 1976 og bygger på BTN)
- e) Varefortegnelse for Utenrikshandelen (omtalt særskilt seinere)
- f) Varenomenklaturen i Industristatistikken
- g) NIMEXE er EF's varenomenklatur for utenrikshandelen, mens NIPRO er EF's nomenklatur for varens opprinnelsessektor (EF's versjon av IGGS). EF-landene har også en varenomenklatur til (CST) som brukes i internasjonal handel
- h) Standard for gruppering av varer i transportstatistikk (CTSE - utarbeidd av FN). Som en "condensed list" foreligger CTSE som en felles nordisk varefortegnelse. Den brukes i samferdselsstatistikkene
- i) NST er EF-landenes nomenklatur for transportstatistikk

3. Klassifisering av produksjonsenheter i bedrifter og foretak

- a) FN's næringsstandard (ISIC, Rev.2, International Standard Industrial Classification of All Economic Activities)
- b) Standard for næringsgruppering (SN - omtales særskilt)
- c) Nordisk standard for næringsgruppering (bare for næringene industri og bergverksdrift)
- d) NACE er EF's næringsstandard
- e) CCIO (Classification of Commodities by Industrial Origin) er FN's nøkkel mellom ISIC og SITC. Den siste er fra 1971 og vil ventelig bli avløst av ICGS (se 2 b)

4. Klassifisering av individer, husholdninger mv.
 - a) FN's yrkesklassifisering (ISCO 1968, International Standard Classification of Occupations)
 - b) Nordisk yrkesklassifisering (NYK; en norsk utgave av NYK er utarbeidd av Arbeidsdirektoratet og utgitt i 1965). Standarden bygger på 1958-utgaven av ISCO (se pkt. 4 a)
 - c) WHO's sykdomsklassifisering (ICD - International Statistical Classification of Diseases)
 - d) Standard for gruppering av sykdommer - skader - dødsårsaker i offentlig norsk statistikk (omtales særskilt seinere)
 - e) Standard for utdanningsgruppering (omtales særskilt seinere)
5. Klassifisering av finans- og realobjekter
 - a) FN's standard på dette området er den standardiseringen som er gitt i SNA
 - b) Norsk standard klassifisering er gitt i de årlige kredittmarkedstatistikk-publikasjonene
6. Regionale standarder
 - a) Standard for handelsområder (omtales særskilt)
 - b) Standard for kommuneklassifisering (omtales særskilt)

B. Standard for næringsgruppering

Standard for næringsgruppering gjennomgikk en fullstendig revisjon omkring 1970. Første utgave av denne standarden ble utgitt av Statistisk Sentralbyrå i 1954. Den andre utgaven ble utgitt i serien Statistisk Sentralbyrås Håndbøker (nr. 9) i 1960. Den reviderte standarden kom første gang i 1972, og seinere (1976 og 1978) har den blitt ajourført, først og fremst på grunn av oljevirkksomheten i Nordsjøen og endringer i skole- og helsesystemet. Standarden er utgitt i serien Håndbøker og heter: Standard for næringsgruppering, forkortet SN.

Standard for næringsgruppering er utarbeidd i Statistisk Sentralbyrå på grunnlag av den seineste utgaven (fra 1968) av International Standard Industrial Classification of All Economic Activities eller ISIC som den forkortes. ISIC er utarbeidd under ledelse av FN.

Det er flere formål med en standard for næringsgruppering. Hovedhensikten med standarden er å sørge for at det ved alle næringsgrupperinger så vidt mulig blir brukt samme klassifikasjonsprinsipper. En ønsker å sammenlikne enheter som har felles egenskaper når det gjelder næringsvirksomhet. Hvis slike enheter samles i næringsgrupper, vil dette kunne gi grunnlag for å sammenlikne, vurdere og gjensidig kontrollere opplysningene (data) hentet inn om enhetene selv om opplysningene er hentet inn fra ulike kilder, oppgavegivere eller fra allerede utarbeidd statistikk.

Foruten å systematisere selve næringsgruppeinndelingen må en da også definere enhetene som skal klassifiseres i næringer. Slike enheter er f.eks. bedrifter; foretak og personer.

Standarden er prinsipielt utformet som et forslag til klassifisering av enheter, både enheter som driver økonomisk virksomhet (bedrifter) og andre tilsvarende enheter. Det må være mulig og naturlig å si at næringsvirksomhet utøves av enheter selv om de da ikke driver økonomisk virksomhet (som f.eks. et departement, et direktorat, en humanitær organisasjon osv.).

Klassifikasjonssystemet er et pyramidesystem, og standarden (SN) opererer med følgende 5 inndelingsnivåer:

Næring	1-sifret kode	Tilsvarende ISIC-nivå
Næringsområde	2-sifret kode	" "
Næringshovedgruppe	3-sifret kode	" "
Næringsgruppe	4-sifret kode	" "
Næringsundergruppe	5-sifret kode	Nasjonalt norsk nivå

I standarden står f.eks. 313 Produksjon av drikkevarer. Vi har her en 3-sifret tallkode foran navnet, altså er det en næringshovedgruppe. Denne hovedgruppen er delt i næringsgruppene 3131 Produksjon av brennevin og vin, 3133 Brygging av øl og 3134 Produksjon av mineralvann. Sammen med de andre næringshovedgruppene 311-312 Produksjon av næringsmidler og 314 Produksjon av tobakkvarer danner næringshovedgruppen 313 næringsområdet 31 Produksjon av næringsmidler, drikkevarer og tobakkvarer. Alle næringsområder som har en 2-sifret tallkode hvor 3 er første siffer, utgjør til sammen næring 3 som er Industri.

I enkelte tilfelle hvor gruppeinndelingen i ISIC bare går ned til 2. eller 3. siffer, er også 3. respektive 4. nivå utnyttet til en mer detaljert norsk inndeling. Bortsett fra slike tilfelle, er næringsgruppeinndelingen på de 4 første nivåer med meget få unntak helt identisk med inndelingen på de tilsvarende nivåer i ISIC.

En standard for næringsgruppering lar seg ikke bygge opp på grunnlag av bare ett klassifikasjonsprinsipp. Hovedprinsippet som har vært fulgt er at foretak, bedrifter, offentlige og private organisasjoner og institusjoner og andre statistiske enheter er gruppert (klassifisert) etter arten av den virksomhet disse enhetene hovedsakelig driver. Den virksomhet som drives er enten framstilling og/eller omsetning av varer og yting av tjenester.

Standarden er altså hva en kan kalle en generell aktivitetsstandard, og den tar f.eks. ikke sikte på å klassifisere etter produksjonsmetode.

Avgrensingen av gruppene faller som nevnt ned til et visst nivå sammen med ISIC's gruppering. Ved den videre gruppeinndelingen fra ISIC's laveste nivå har en klassifisert ut ifra det siktemål at de mer detaljerte gruppene skulle være mest mulig homogene med hensyn til: a) den tekniske organiseringen av vare- og tjenesteproduksjonen, b) til de produserte varenes og tjenestenes art og c) anvendelsesområde. Det er ved utarbeidelsen av den nye standarden tatt sikte på at de enkelte gruppene skulle ha en forholdsvis høy dekningsgrad og spesialiseringsgrad. Høy dekningsgrad for en gruppe vil si at det i vedkommende gruppe foregår den vesentlige delen av den norske produksjonen av de varer og tjenester som karakteriserer gruppen. I standarden er det angitt under hver næringsgruppe hvilke varer som er karakteristiske for gruppen, f.eks. er melk en karakteristisk vare for næringsundergruppen Meierier. Høy spesialiseringsgrad vil si at gruppen hovedsakelig produserer de varer som karakteriserer gruppen og ikke hovedsakelig andre varer. Fordi mange statistiske enheter (bedrifter, foretak osv.) har en forholdsvis blandet produksjon, vil deknings- og spesialiseringsgraden for de fleste gruppene være mindre enn 100 prosent. I NOS Industristatistikk er det i en tabell gitt tall for deknings- og spesialiseringsgrad (homogenitetstall).

I tillegg til de klassifikasjonsprinsipper som er nevnt ovenfor, har en ved utarbeidelsen av den siste standarden, også tatt hensyn til ISIC-gruppenes størrelse (målt ved sysselsetting, produksjons- eller bearbeidingsverdi etc.). Det har også, ved utarbeidelsen av den siste standarden, vært lagt vekt på å knytte en så god som mulig forbindelse med tidligere utgaver av standarden.

Siden Standard for næringsgruppering forutsetter at enhver næringsgruppering primært bygger på enheter som vi kaller "bedrifter" er det av største betydning at begrepet bedrift blir oppfattet likt over alt. Det er ellers ikke mulig å jamføre resultatene av forskjellige statistiske undersøkelser (jfr. kravet til sammenliknbarhet).

Standarden skiller også mellom "bedrift" og "foretak". Et "foretak" defineres som en eller flere bedrifter som drives av samme eier (den juridiske person). For foretaket kan det i prinsippet alltid stilles opp en fullstendig status og en slags vinnings- og tapskonto (resultatkonto).

Foretaket vil som regel falle sammen med det som ofte betegnes "firma" eller selskap, men av og til også (dessverre) det som i dagligtalen betegnes med bedrift. Når en i aviser, tidsskrifter o.l. ser ordet "bedrift", er ikke alltid enheten ordet refererer seg til det samme som standarden har definert.

Bedrift er i standarden definert som en lokalt avgrenset funksjonell enhet hvor det hovedsakelig drives aktiviteter som faller innenfor en bestemt næringsgruppe.

En bedrift kan derfor være det vi i dagligtale omtaler som en fabrikk, en produksjonsavdeling, en gruve, en butikk osv.

Hvis f.eks. et aksjeselskap driver virksomheter på lokalt atskilte områder, vil de lokalt atskilte delene regnes som særskilte bedrifter.

I et flerbedriftsforetak, dvs. foretaket består av flere enn en bedrift, kalles en av bedriftene for hovedbedriften. Det er da den bedriften av samtlige bedrifter i foretaket som produserer de produkter som gir størst bearbeidingsverdi eller bruttofortjeneste. (Litt upresist kan vi si at bearbeidingsverdien er det samme som brutto produksjonsverdien minus vareinnsatsen.)

Hvis det er tvil om hvorvidt en bedrift eller et foretak skal plasseres i en bestemt næring, er det bearbeidingsverdien som skal avgjøre plasseringen. Bedriften klassifiseres etter det produkt eller den gruppe av produkter som har høyest bearbeidingsverdi. Foretaket skal i prinsippet klassifiseres i den gruppe hvor hovedbedriften er plassert. Hvis et foretak består av mange bedrifter og at noen eller alle de andre bedriftene utenom hovedbedriften hører hjemme i samme næring og at disse andre bedriftene til sammen har høyere bearbeidingsverdi enn hovedbedriften, bør foretaket plasseres i disse bedriftenes næring.

Den nye standarden har følgende næringer (jfr. klassifikasjonssystemet, dvs. 1-sifret nivå):

1. Jordbruk, skogbruk, fiske og fangst
2. Bergverksdrift
3. Industri
4. Kraft- og vannforsyning
5. Bygge- og anleggsvirksomhet
6. Varehandel, hotell- og restaurantvirksomhet
7. Transport, lagring, post og telekommunikasjoner
8. Bank- og finansieringsvirksomhet, forsikringsvirksomhet, eiendomsdrift og forretningsmessig tjenesteyting
9. Offentlig, sosial og privat tjenesteyting

C. Vareklassifisering; Brussel-nomenklaturen og FN's standard

I offisiell norsk statistikk er to standarder i bruk for vareklassifisering, dvs. standardisering av varegrupper. Enhetene som vi ønsker å klassifisere i grupper, er de enkelte varer. Kjennermerker som vi beskriver enhetene (dvs. varene) med, er varens mengde, varens verdi, varens produsent (land og/eller næring) osv. For alle varer er også varens mengde standardisert, dvs. at enhetene (vi klassifiserer) er en bestemt mengde (kg, stk., osv.) av en vare.

Byrået utarbeider flere statistikker som bruker de standardiserte varegrupperingene. Først og fremst Utenrikshandelsstatistikken og Industristatistikken.

De to standarder heter FN-standard (SITC) og Brussel-nomenklaturen (BTN). Fra og med 1976 er betegnelsen på Brussel-nomenklaturen endret til Customs Co-operation Council Nomenclature (CCCN), men det vil nok ta en tid før den nye benevnelsen er tatt i bruk over alt.

FN's standard SITC (Standard International Trade Classification) ble tatt i bruk av Byrået i 1953. Denne standarden ble revidert i 1961 og i 1975. Den siste reviderte standard er både av Byrået og Tollvesenet vedtatt gjennomført fra 1. januar 1976. Standarden betegnes nå SITC - Rev. 2.

Fra 1. januar 1959 ble den nye tolltariffen (dvs. en tariff for beregning av toll) som bygger på Brussel-nomenklaturen, tatt i bruk av Byrået for statistikken over utenrikshandelen. Denne statistikken utgir også tall for varer gruppert etter SITC - Rev. 2. Hvert år gir Byrået ut et hefte som (i 1980) heter Statistisk varefortegnelse for utenrikshandelen i serien Norges offisielle statistikk (NOS). Dette heftet er samtidig Håndbok nr. 26 (som altså da blir revidert hvert år).

Tolltariffen egner seg imidlertid ikke uten videre som varefortegnelse for industribedrifters oppgave over vareproduksjon og vareforbruk. Den er på visse felter altfor spesifisert, mens den andre steder er for summarisk. Det er derfor utarbeidd en særskilt varefortegnelse for industristatistikken. Varefortegnelsen er også utgitt i serien Statistisk Sentralbyrås Håndbøker (nr. 19) og heter Vare-nomenklatur for industristatistikken. Den utgis nå årlig som et vedlegg til Industristatistikk-publikasjonen. Industristatistikkens varefortegnelse er så langt det er funnet hensiktsmessig, samordnet med tolltariffen. Det er nå i gang et fellesprosjekt mellom Utenrikshandelsstatistikken og Industri-statistikken med sikte på større ensartethet i varespesifikasjonene.

Før vi går over til å se på hvordan sammenhengen er mellom Brussel-nomenklaturen (som toll-tariffen bygger på) og FN-standard, vil vi se mer detaljert på de to standarder hver for seg.

Brussel-nomenklaturen (BTN eller nå CCCN) har 21 avsnitt, 99 kapitler og ca. 1 100 basisposisjoner. Prinsippet bak kapittel- og posisjonsinndelingen i BTN er kravet om at varene skal deles i grupper etter varer som består av samme råstoff. Dvs. at i de samme varegruppene (BTN-posisjoner) kommer både råvarer, halvfabrikata og ferdigvarer. På grunnlag av denne nomenklaturen er det laget en statistisk varefortegnelse. For hver enkelt vare i varefortegnelsen er det fastsatt et syvsifret kodenummer. Dette kodenummer er bygd opp på følgende måte:

Første siffer:	B	= BTN-avsnitt
De to første siffer:	BB	= BTN-kapittel
De fire første siffer:	BBBB	= BTN-posisjon
De seks første siffer:	BBBB.NN	= Nordisk varepost
Alle syv sifrene:	BBBB.NNn	= Nasjonalt varenummer

Ved oppdeling av BTN-posisjonene i vareposter og varenummer er det tatt hensyn til:

- 1) at hvert varenummer (BBBB.NNn) helst skal kunne plasseres under en bestemt varepost (item) i SITC - Rev. 2 (se seinere),
- 2) de nordiske tolltariffer og andre nordiske fellesinteresser,
- 3) den norske tolltariffen,
- 4) spesiell norsk statistisk interesse.

Selv om en altså har forsøkt å ta statistiske hensyn på den mest detaljerte inndelingen, er BTN først og fremst beregnet til tollbruk. En klassifisering av varer etter BTN gir oss ikke de samme analysemuligheter som SITC - Rev. 2. FN-standard er nemlig utformet til statistisk bruk. I SITC - Rev. 2 er graden av bearbeidelse av varene hovedprinsippet som er lagt til grunn for grupperingen av varene. Det vil igjen si at råvarer, halvfabrikata og ferdigvarer kommer i forskjellige SITC-grupper. Ofte er det av større interesse i økonomisk-statistisk analyse å se på i hvor stor grad varene er bearbeidd.

I SITC - Rev. 2 er det følgende grupperinger og tallkoder:

S	= varesektorer (sections), i alt	10 varesektorer
SS	= hovedgrupper (divisions), i alt	63 hovedgrupper
SSS	= grupper (groups), i alt	233 grupper
SSS-S	= undergrupper (subgroups), i alt	786 undergrupper
SSS-SS	= vareposter (items), i alt	1 924 vareposter

Den mest detaljerte inndelingen angis altså med 5-sifrete koder. I tillegg inkluderer noen koder, i parentes, en liten bokstavkode.

Sammenhengen mellom BTN og SITC - Rev. 2 kommer tydeligere fram ved et eksempel:

Kapittel 24 i BTN heter Tobakk, og for dette kapitlets vedkommende er sammenhengen med SITC - Rev. følgende:

BTN-nr.		SITC - Rev. 2 nr.
	Kapittel 24. Tobakk	
24.01	Tobakk, ubearbeidd; avfall av tobakk	
2401.100	tobakk, ubearbeidd	121-00
2402.200	avfall av tobakk	"
24.02	Tobakk, bearbeidd; tobakkekstrakter og tobakkessenser	
2402.110	sigarer og sigarillos med vekt over 3 g pr. stk.	122-10
2402.120	sigarer og sigarillos med vekt 3 g eller under pr. stk.	"
2402.200	sigaretter	122-20
2402.300	røyketobakk	122-30
2402.500	annen tobakk	"
2402.600	tobakkekstrakter og -essenser	"

Vi ser at det er samsvar mellom det syvsifrete BTN-nummer (dvs. det nasjonale varenummer) og det femsifrete SITC - Rev. 2 nummer (dvs. varepost). Sammenhengen er ikke entydig fordi det til et SITC - Rev. 2 nummer kan svare flere BTN-nummer.

Forskjellen mellom de to standardenes (nomenklaturenes) klassifiseringsprinsipper kan belyses ved følgende eksempel: I Brussel-nomenklaturen er det naturlig å plassere i samme kapittel alle varer som i første rekke består av jern og stål (kapittel 73). Derimot kan man i SITC-standardens finne produkter som består av jern og stål på flere steder innen varesektorene 2, 6 og 8. Det er en tendens i SITC-standardens til å gruppere sammen homogene varer og innen disse gruppene klassifisere etter råvareinnholdet. I CCCN gjør man akkurat det motsatte.

Industristatistikkens varenomenklatur følger BTN's nummerkode i stor utstrekning; imidlertid er det av og til visse avvik. F.eks. er det nasjonale BTN-nummer 2402.500 (annen tobakk) i industristatistikken delt i to, nemlig 2402.501 (skråtobakk) og 2402.509 (annen tobakk). På industristatistikk-skjemaene er det foran selve varenummeret (prefiks) tilføyet et ettsifret kodetall som angir om vedkommende vare er nyttet som råstoff eller som produksjon. Kode 1 for råstoff og kode 0 for produksjon, men det finnes også andre tallprefiks.

Den årlige publikasjonen Utenrikshandel er delt i to hefter, hvor hefte I inneholder opplysninger om innførsel og utførsel av varer etter Brussel-nomenklaturens varegruppering, mens hefte II følger den internasjonale handelsstatistiske nomenklatur (FN's SITC - Rev. 2). I publikasjonen er det også en nøkkel for sammenhengen mellom de to nomenklaturene.

Produksjon av de enkelte varer er i den årlige industristatistikkpublikasjonen gitt syvsifret varenummer. I denne statistikken er det avmerket hvis det er avvik mellom tolltariffen og industristatistikkens gruppering. Avmerkingen er foretatt ved hjelp av en X etter varenummeret.

D. Standard for utdanningsgruppering

Utdanning er, som det heter i denne standarden¹⁾(NUS), i vid forstand en fellesbetegnelse på formidling av ferdigheter og kunnskaper. Utdanningssystemet defineres i standarden slik at det omfatter all fast organisert virksomhet spesielt innrettet på systematisk formidling av ferdigheter, kunnskaper og forståelse: Med en utdanningsaktivitet forstås en gruppe utdanningstilbud som er like med hensyn til opptakskrav, varighet, fagkrets, klasseinndeling, heltids-/deltidsutdanning og utdanningsinstitusjon.

Formålet med standarden er å klassifisere de delaktiviteter, utdanningsaktiviteter, som utdanningssystemet består av. Enhetene en ønsker å klassifisere er altså utdanningsaktivitet. Til å karakterisere disse enhetene er brukt de kjennemerker som er nevnt i definisjonen: a) obligatorisk eller ikke-obligatorisk utdanning, b) heltids- eller deltidsutdanning, c) klasseinndeling, d) opptakskrav, e) varighet, f) fagkrets og g) utdanningsinstitusjon.

Når en har fått delt utdanningssystemet opp i standardiserte aktiviteter, ønsker en av flere grunner å se hvor mange personer som enten er i en bestemt utdanningssituasjon (dvs. kan henføres til en bestemt aktivitet) og hvor mange som i en periode har gjennomført denne utdanning. Selvfølgelig er en ikke bare interessert i antall personer, men også i visse kjennemerker hos disse personene, som f.eks. kjønn, alder, bosted, familiebakgrunn, forutdanning og eksamensresultater. Den enkelte person blir derfor på samme måte som utdanningsaktivitet oppfattet som en (grunnleggende) enhet i utdanningsstatistikken.

Klassifikasjonsprinsippet som standarden bygger på, er at enheten utdanningsaktivitet er gruppert etter kombinasjoner av kjennemerkeverdiene for opptakskrav, varighet og fagkrets. En gruppe utdanningsaktiviteter, kjennemerket ved et bestemt sett verdier på disse kjennemerkene, kalles for en enkelt-utdanning.

Også denne standarden er bygget opp som et pyramidisk klassifikasjonssystem. Enkeltutdanningene, som altså består av en gruppe utdanningsaktiviteter, er først ordnet i grove grupper, utdanningsnivåer, etter verdien på kjennemerkene opptakskrav og varighet. Hvert utdanningsnivå er delt opp i fagfelt, og hvert fagfelt er delt opp i faggrupper. De to siste oppdelingene er begge foretatt på grunnlag av verdier på kjennemerket fagkrets. Faggruppene er endelig til slutt delt opp i utdanningsgrupper.

1) Statistisk Sentralbyrås Håndbøker nr. 28: Standard for utdanningsgruppering i offentlig norsk statistikk.

Det er i alt 9 utdanningsnivåer, og hvert nivå er delt i 9 fagfelt og hvert fagfelt i 9 faggrupper. Fagfeltene er like innenfor alle nivåene, men faggruppene kan variere noe fra fagfelt til fagfelt.

I utdanningsstandarden har Byråskolens kurs III tallkoden 4438.3, dette vil si at Byråskolen er gruppert under nivå 4 (utdanning på gymnasnivå) fagfelt 4 (Administrasjon og økonomi), faggruppe 3 (kontorfag) og utdanningsgruppe 8 (spesialutdanning for offentlig ansatte kontorfunksjonærer). Utdanningsgruppe 8 består av i alt 4 enkeltutdanninger: 4438.1 Arkivlederutdanning, 4438.2 Kontorelevutdanning for nyansatte i staten, 4438.3 Ett-årig saksbehandlerkurs og 4438.4 Byråskolen, kurs III.

E. Standard for handelsområder¹⁾

Det er flere formål med å inndeile landet i geografiske områder. Et formål kan være av administrativ art, og da er landet delt opp i administrative enheter. Et hovedprinsipp ved alle administrative inndelinger er at de bygger på den enkelte kommune som enhet, mens de har varierende høyere enheter i ett eller flere trinn, men alle de høyere inndelingene følger alltid kommunegrenser.

Et annet formål med geografiske inndelinger er å gi et grunnlag for næringsøkonomiske analyser. En ønsker å sammenlikne næringslivet i et område med næringslivet i et annet. De enheter som en vil beskrive næringslivet ved hjelp av, innhentes det statistiske opplysninger om fra flere kilder, og ved geografiske sammenlikninger er det da viktig at de geografiske inndelingene er standardisert.

Klassifikasjonsprinsippet har vært at landet skulle delles opp i funksjonelle områder. Et område er funksjonelt når de enkelte deler av det er innbyrdes avhengige og er bundet sammen på en eller annen måte, f.eks. ved utveksling av varer og tjenester. Vi har et særpreget tilfelle av et funksjonelt område når det innen et område foregår en orientering mot et sentralt sted.

Hovedkriteriene ved fastleggelsen av områdene var den innenlandske vareomsetningen, bosetningen og kommunikasjonsforholdene.

Klassifikasjonssystemet er et pyramidesystem hvor landet er delt opp på følgende måte:

G = Handelsfelt, i alt er det 4 felt

GG = Handelsområde, hvert felt er delt i områder, i alt 23 områder

GGG = Handelsdistrikter, hvert område er delt i distrikter, i alt 105 distrikter, som er satt sammen av et varierende antall herreds- og bykommuner

Kodetall 342 står da for felt 3 (Vestre handelsfelt), område 4 (Sunnmøre handelsområde, som ligger innenfor felt 3) og distrikt 2 (Volda/Ørsta handelsdistrikt, som ligger innenfor område 4).

Handelsdistriktsinndelingen ble siste gang revidert i samsvar med kommuneinndelingen pr. 1. januar 1972. Dette var bare en korreksjon for endringer i kommunegrenser siden den siste mer omfattende revisjon av inndelingen som ble foretatt i 1966. Den nåværende handelsdistriktsinndelingen er derfor tilpasset samfunnsforholdene (bl.a. fylkenes planleggingsområder) for 10 år siden, og det er derfor ut fra dette et behov for revisjon, men samtidig er det ut fra prinsippet om minst mulig forandring i standardene slik at de kan tjene sin hensikt som grunnlag for sammenlikninger over et lengre tidsrom, ønskelig at standarden ikke revideres. Byrådet bestemte i 1977 at standarden ikke ville bli revidert før 1980.

F. Standard for gruppering av sykdommer - skader - dødsårsaker

Standarden er utgitt i serien Statistisk Sentralbyrås Håndbøker (nr. 24). Det er den åttende revisjon av den internasjonale statistiske standard for sykdommer, skader og dødsårsaker, godkjent av Verdens helseorganisasjon, som den norske utgaven bygger på. Klassifikasjonene ble tatt i bruk i offentlig norsk statistikk i 1969.

1) Omtalt i Statistisk Sentralbyrås Håndbøker nr. 13: Standard for handelsområder.

Formålet med standarden er å klassifisere sykdommer, skader og dødsårsaker i grupper som tilfredsstiller det statistiske behov til tre ulike brukerkategorier:

1. Statistisk Sentralbyrå (til dødsårsaksstatistikken og annen offisiell helsestatistikk)
2. Rikstrygdeverket og trygdekassen (til morbiditetsstatistikk, yrkesskade- og sykdomsstatistikk, og statistikk over årsaker til uførepensjon etter folketrygdloven)
3. Sykehusene og andre medisinske institusjoner (til føring av diagnosekartoteker og til eventuell morbiditetsstatistikk)

Det er et statistisk klassifikasjonsprinsipp som er lagt til grunn for standarden. En tar sikte på å gruppere de ulike diagnostiske termini¹⁾ som brukes på forskjellige plan i vår medisinske virksomhet.

Denne standarden er utarbeidd på nordisk basis som en femsifret tilpasning til den internasjonale standarden, som omfatter de tre første sifrene. Klassifikasjonssystemet kan knapt kalles pyramidisk: Det første siffer er skrevet ved hjelp av romerske tall, mens resten er angitt ved arabiske tall. Det romerske tallet står for kapittel (i alt 17 stykker) og er unødvendig nær en diagnose skal plasseres, fordi det første siffer (dvs. de to første pga. 17 hovedgrupper) ikke varierer fra kapittel til kapittel.

Verdens helseorganisasjon har lenge arbeidd med en 9. revisjon. De nordiske land har sagt ifra at de ikke ønsker endringer allerede.

G. Andre statistiske standarder

Standard for kommunklassifisering²⁾

Standard for kommunklassifisering (SK) er først og fremst beregnet for bruk i offentlig norsk statistikk, men den er også nyttet av forskere.

Kommunene er klassifisert i en såkalt grunnklasse på grunnlag av tre kriterier (tre-leddet grunnkode): Første ledd representerer næringsstrukturen, annet ledd representerer bosetningstettheten, og tredje ledd representerer kommunens såkalte sentralitet (dvs. kommunens geografiske beliggenhet i forhold til sentra (dvs. byer o.l.). På basis av første og tredje ledd i grunnkoden er kommunene gruppert i ni kommunetyper.

Eks.:

Kommune nr.	Navn	Grunnkode	Type
0723	Tjøme	TT - 3 - 2A	8

hvor TT (første ledd) står for tjenesteyting, 3 i annet ledd står for bosetningstettheten på Tjøme, og 2A sier hvor sentral Tjøme er. Typekoden (8) sier at Tjøme er en av de såkalte "Øvrige blandede tjenesteytings- og industrikommuner".

Standarden ble første gang publisert som en foreløpig standard i 1972, og den endelige, nåværende standarden kom i 1975.

Inndeling av landet i jordbruksområder

En slik inndeling har vært brukt i Byrået helt fra jordbrukstellingen i 1929. Den er foreldet, og ble ikke brukt ved tellingen i 1969.

Standard for husdyrsykdommer

Dette er en standard som Veterinærdirektoratet har størst interesse av (brukes bl.a. i kontrollvirksomhet), men som også brukes i statistikkproduksjonen (NOS Veterinærstatistikk).

1) Diagnostiske termini vil si det samme som de ord og uttrykk som stikkordmessig angir en bestemt diagnose (karakteriserer den patologiske tilstand).

2) Statistisk Sentralbyrås Håndbøker nr. 35.

Standard for varegruppering i transportstatistikken

I samferdselsstatistikken følger Byrået den såkalte "Condensed list", som er en felles nordisk varefortegnelse på grunnlag av CTSE (Den europeiske nomenklatur for gruppering av varer i transportstatistikker). Blant andre Norges Statsbaner gjør bruk av denne standarden i sin virksomhet. Det er egenskaper ved varene som er relevante i transportsammenheng som har dannet grunnlag for varegruppene. I alt er det 20 grupper i CTSE (korn, frisk frukt og grønnsaker,, tjære av kull og naturgass,, metaller, arbeider av metaller etc.). Den nordiske utgaven gir en videre oppdeling av disse gruppene (f.eks. er korn delt i hvete og annet).

4. OM SENTRALE STATISTISKE REGISTRE

A. Hva er et register?

For analyse av mange samfunnsvitenskapelige problemstillinger vil det være av stor betydning at en kan knytte sammen primærmateriale for de enkelte enheter fra forskjellige kilder og tidspunkter. Skal dette lykkes, må tre betingelser være oppfylt:

- 1) Enhetene må være likt definert i de ulike kildene (standardisering)
- 2) Det må opprettes og holdes à jour registre over enhetene hvor hver enkelt er entydig identifiserbar
- 3) Observasjonene må samles, organiseres og lagres slik at de uten for store kostnader kan kombineres og utnyttes i analyser

Vi har foreløpig bare sett på den første betingelsen. I dette kapitlet vil vi ta for oss registre, først og fremst de to sentrale registrene, som finnes i Byrået: Personregisteret og Bedrifts- og foretaksregisteret.

Med et register (som på engelsk ofte betegnes som "file") forstår vi den totale mengde kjennemerkeverdier som er lagret på en systematisk måte. Registeret bygges opp av et antall enheter, eller poster (på engelsk "records"), hvor hver post inneholder et visst antall felt, eller kjennemerker. Hvert felt består av et visst antall tegn som til sammen gir oss kjennemerkeverdien.

Denne formelle definisjon er lang og inneholder begreper og formuleringer som gjør den lite tilgjengelig. Vi skal derfor bruke følgende definisjon:

Et register er en "fortegnelse" over en bestand av enheter som på et gitt tidspunkt tilfredsstiller en gitt definisjon, dvs. har visse felles egenskaper.

Et register er altså en fortegnelse over enhetene i en nærmere definert masse.

F.eks. krever et register over personbiler av modell 1972 registrert i Trøndelag i 1973, at de enheter som kan inntas i registeret har verdiene personbil, 1972, Nord- eller Sør-Trøndelag og 1973 på de fire kjennemerkene transportmiddel, modell, registreringsdistrikt og registreringsår.

I et register må det være registrert kjennemerker som identifiserer de enkelte enhetene som registeret består av. Identifikasjonen kan være av intern eller ekstern karakter. Intern identifikasjon vil si et nummersystem som først og fremst brukes innen registersystemet. Ekstern identifikasjon brukes når enhetene skal identifiseres utenfor registeret, f.eks. navn og adresse i et personregister. Et register kan inneholde flere kjennemerker enn identifikasjonskjennemerker. I en del litteratur blir et slikt register omtalt som et dataarkiv, mens det i annen litteratur ikke skjelnes mellom registre og dataarkiv. Å skjelne mellom register og dataarkiv etter dette kriteriet har vist seg å være vanskelig i praksis, og vi vil holde oss til benevnelsen register. Begrepet dataarkiv vil bli brukt om et "sted" hvor vi oppbevarer data. Det kan f.eks. være et rom hvor vi oppbevarer skjema, hullkort, magnetbånd o.l., men begrepet kan også brukes i mer abstrakt betydning. Således er det andre egenskaper ved en datamasse enn antall kjennemerker som bestemmer om en datamasse skal betegnes som et register eller ikke. Se definisjonen.

B. Hva brukes registre til?

En totaltelling kan nok i de fleste tilfelle gjennomføres uten bruk av registre, men det sier seg selv at det forenkler arbeidet dersom en fortegnelse over samtlige enheter i massen er tilgjengelig. Skal det f.eks. sendes brev til samtlige enheter (f.eks. personer), må adressene finnes i et register. I eksperimentalt statistikk har man ikke det samme behov for registre, der kommer den ene enheten fram etter den andre. Det er behovet for navne- og adresselister som er den viktigste årsaken til at registre er opprettet og brukes i forbindelse med totaltellinger.

En utvalgsundersøkelse må ofte bygge på et register, men det er ikke nødvendig i alle slike undersøkelser. Ved stikkprøveundersøkelser, som altså er et annet navn på utvalgsundersøkelser, kan man plukke ut både et tilfeldig antall enheter og tilfeldige enheter av den totale masse. Dreier det seg om en bilkontroll f.eks., plukkes det ut bare noen biler nå og da som kontrolleres. Hensikten med en slik bilkontroll er bl.a. å få bilde av den tekniske standarden i hele bilparken. Formålet med utvalgsundersøkelser er å kunne si noe om hele massen ut ifra opplysninger om utvalget.

Når vi foretar en utvalgsundersøkelse, vil de utsagn vi kommer med om enhetene i hele massen bygge på de målingene vi foretok av enhetene i utvalget. Vi ønsker på den ene siden at utvalget skal være så lite som mulig fordi kostnadene ved undersøkelsen øker med antall enheter, og på den andre siden vil et større utvalg kunne gi muligheter for mer nøyaktige utsagn om hele massen enn et lite utvalg. Det er ikke bare utvalgets størrelse som har noe å si for nøyaktigheten eller presisjonen, men også på hvilken måte utvalget trekkes, dvs. hvordan de enkelte enhetene i utvalget blir plukket ut av den totale massen. Og skal vi på en bestemt måte trekke ut et visst antall enheter av en masse, må vi som oftest ha et register. Behovet for register vil avhenge av måten utvalget trekkes på. I Praktisk statistikk av E. v. Hofsten¹⁾ er det omtalt i alt 7 måter eller metoder å trekke et utvalg på.

Hensikten med å trekke utvalg på bestemte måter er at vi med større sikkerhet kan uttale oss om hele massen på grunnlag av opplysninger om bare utvalget. Det er da av stor viktighet at registeret setter oss i stand til å trekke utvalget slik at nøyaktigheten blir størst mulig innenfor den kostnadsramme vi som oftest må sette.

Foruten at registeret brukes både ved total- og utvalgsundersøkelser, brukes registre også ved datainnsamling fra såkalte administrative prosesser. Dette betyr at statistiske data som f.eks. Byrået skal innhente fås fra en administrativ rutine (dvs. en annen offentlig etat), og ikke direkte fra den enkelte enhet. En administrativ rutine er f.eks. selvangivelsene. Byrået er av statistiske grunner interessert i opplysningene som finnes på selvangivelsene, og Byrået "får" selvangivelsesopplysningene fra likningskontorene og ikke fra de personer etc. som leverer selvangivelser. Hvis Byrået ønsker å koble sammen de opplysningene som fås inn, f.eks. via selvangivelsene og via forbruksundersøkelsen, må enhetene i de to massene enten være like eller det må finnes muligheter for å bygge opp enhetene i den ene massen ved hjelp av enhetene i den andre massen.

Datainnsamling direkte fra de enkelte oppgavegiverne er kostbart, men datainnsamling nærmest som et biprodukt ved administrative prosesser kan være meget billig for Byrået. I enkelte tilfeller vil datainnsamling via administrative prosesser for en hel masse være like billig som en utvalgsundersøkelse for statistikkprodusenten.

Å knytte sammen ulike primærmaterialer med samme enhetsdefinisjon, kan danne grunnlag for ny statistikk og utvidede og mer verdifulle undersøkelser og analyser. Et fullstendig register over enhetene vil kunne være til stor hjelp i et slikt prosjekt.

O p p s u m m e r i n g

Register brukes til:

1. Administrative formål
2. Statistiske formål
 - a. Innhenting av oppgaver fra enhetene
 - b. Trekking av utvalg
 - c. Direkte grunnlag for statistikk (situasjon, endringsmeldinger)
 - d. Sammenkobling av data fra ulike kilder

1) I kapitlet Urvaldsmetoder, side 49-58. Dessuten behandlet mer utførlig i faget Matematisk statistikk.

C. Prinsipper for oppbygging og ajourhold av registre

Det er vanskelig å gi generelle prinsipper for hvordan et register skal oppbygges og ajourholdes. Eventuelle prinsipper må utledes av de formål registeret skal tjene. De ulike måter å bruke registeret på, som er nevnt ovenfor, fører til at et register for å være brukbart må tilfredsstille de fleste av følgende krav:

- 1) Registeret må være fullstendig, dvs. alle enheter som inngår i massen må være med. Dette må gjelde på ethvert tidspunkt, dvs. registeret må kunne holdes à jour
- 2) Registeret må bare inneholde enheter som inngår i massen og ikke enheter som ikke tilhører massen
- 3) De enkelte enhetene må bare inngå en gang i registeret
- 4) Opplysningene som registeret inneholder om de enkelte enhetene må kunne holdes à jour

Når det gjelder ajourholding, kan den enten være løpende eller den kan skje på visse fastlagte datoer (f.eks. rett før det skal brukes hvis det f.eks. bare brukes noen få ganger i året).

For at registeret skal kunne holdes à jour med hensyn til tilgang og avgang av enheter, enten det gjelder løpende ajourhold eller ikke, må hver enhet i registeret (dvs. i massen) være utstyrt med et entydig og permanent identifikasjonsnummer eller annen identifikasjon.

D. De sentrale registre i Byrået

De viktigste sentrale registre i Byrået er personregisteret og bedrifts- og foretaksregisteret. Samtlige enheter i disse to registre er identifisert eksternt ved navn og adresse og internt ved et permanent identifikasjonsnummer. Alle andre registre og dataarkiver i Byrået bruker identifikasjonsnumrene i disse to basisregistrene. Identifikasjonsnummeret er også tatt i bruk utenfor Byrået, særlig fødselsnummeret har fått utbredelse.

1. Det sentrale personregister

I Norge fikk vi i 1946 en lov om at alle kommuner var pliktige til å registrere alle bosatte personer. Sentralkontoret for folkeregistrering ble opprettet for å lage forskrifter og å føre kontroll med de lokale registre. Dette kontoret ble lagt til Statistisk Sentralbyrå. Byrået fikk også ansvaret for å etablere et sentralt personregister i 1964. Den umiddelbare bakgrunn var ønsket, særlig fra næringslivet, om å erstatte en rekke forskjellige løpenummersystemer over personer som da var i bruk i flere offentlige etater, med et felles identifikasjonsnummersystem som skulle innføres i alle offentlige etater og også i private institusjoner. Registeret skal tjene både administrative og statistiske formål. En primær oppgave også i dag er tildeling av fødselsnummer. Et entydig og permanent fødselsnummer tildeles for nyfødte og innvandrede som et ledd i den løpende folkeregistrering.

Det sentrale personregister omfatter alle personer som er registrert bosatt i Norge. Mer presist: Når et fødselsnummer først er tildelt, vil det aldri bli fjernet fra registeret.

Følgende kjennemerker blir i dag holdt løpende à jour i personregisteret:

- fødselsnummer
- navn
- mellomadresse
- poststed.
- kommune
- ekteskapelig status
- morens fødselsnummer
- farens fødselsnummer
- ektefellens fødselsnummer
- familienummer
- fra-/tilflyttingsland
- statsborgerskap
- flyttedato

Foreldres og ektefelles fødselsnummer omfatter bare fødsler og giftermål som er meldt personregisteret etter 1. oktober 1964. Familieenheten er definert som en ren kjernefamilie, familienummer er hovedpersonens fødselsnummer. Familiesituasjonen pr. 1. november 1970 ble overført til personregisteret når folketellingsmaterialet var kontrollert og overført til bånd. Endringer i familienummer meldes løpende til registeret fra og med 1. november 1970.

På grunnlag av utdanningsopplysningene fra folketellingen 1970 og løpende meldinger fra utdanningsinstitusjoner er det etablert et dataarkiv for de enkelte personers utdanningsbakgrunn. Dette registeret kan betraktes som et selvstendig register, men som gjennom fødselsnummer kan knyttes til personregisteret.

I enkelte av de lokale folkeregistrene finnes fremdeles lokale registre stort sett ført på hullkort og som nyttes til administrative formål, bl.a. skattemantall og valgmanntall. Ordningen med lokale hullkortregistre innebærer et omfattende dobbeltarbeid når det gjelder dataoverføring. Nedlegging av de lokale hullkortregistre forutsetter at det sentrale personregister kan tilby de samme tjenester som hullkortregistrene nå gir. I så fall må personregisteret innføre følgende kjennemerker:

- valgkrets
- skolekrets

Valgkrets og skolekrets kan bare holdes à jour via et fast kretsregister. Det arbeides med å etablere et fast kretsregister. Videre innebærer en slik administrativ utnyttelse av personregisteret en ytterligere standardisering, bedre kvalitet og en omfattende meldingskontroll av mellomadresseopplysningene. Ajourholdet av personregisteret er i dag i sin helhet basert på meldinger fra folkeregistrene.

Det offisielle fødselsnummersystem nyttes nå ved en rekke administrative registreringer bl.a. innen folkeregistreringen, skattevesenet, trygdeetater, helsevesenet, militærvesenet, undervisningsinstitusjoner og valgmanntall. Også private institusjoner som banker, forsikringsselskaper og organisasjoner nytter fødselsnummeret som identifikasjonsnummer.

I 1977 fikk vi en lov om personregistre¹⁾ m.m. som gjelder for alle personregistre i både offentlig og privat virksomhet. Alle personregistre, bortsett fra noen få unntak (legeregistre, personalarkiver, o.l.), må ha konsesjon for å kunne opprettes og drives. Konsesjonsmyndigheten er tillagt Datatilsynet, som er et nytt frittstående organ under Justisdepartementet. Datatilsynet skal også informere og gi råd om personvern og datasikring. Spesielt skal Datatilsynet vurdere om de forskjellige personregistrene som får konsesjon også skal få anledning til å bruke personnummeret. Loven er en typisk fullmaktslov og derfor vil loven bli gjenstand for nærmere fortolkninger og presiseringer i tiden framover. Byrådet har helt siden Personregisteret ble opprettet praktisert strenge regler for utenforståendes bruk av folkeregisterdata. Med få unntak blir alle nektet selv opplysninger om navn og adresse fra dette registeret. Visse opplysninger i Personregisteret blir brukt bare til statistiske formål.

2. Bedrifts- og foretaksregisteret²⁾

I samband med bedriftstellingene 1953 ble det etablert et bedrifts- og foretaksregister i Statistisk Sentralbyrå. I 1965 ble det innført permanente identifikasjonsnummer for foretak og bedrifter. Nummerseriene for de to typer av enheter er uavhengige.

Bedrifts- og foretaksregisteret ble etablert for å tjene statistiske formål. Registeret skal sikre ensartet klassifikasjon av enheter når oppgaver fra samme bedrift og foretak inngår i ulike statistikker. Det skal dessuten være grunnlag for å trekke utvalg av bedrifter og tjene som adressefortegnelse ved postundersøkelser, som i dag er den vanlige innsamlingsmetode for data fra økonomiske enheter.

Bedriftsenheten er definert for statistiske formål i overensstemmelse med standard for næringsgruppering (se side 18). Foretaksenheten er også definert som i næringsstandarden, og den faller ofte, men ikke alltid, sammen med avgiftspliktig enhet i avgiftsmanntallet, skattepliktig enhet i skattemantallet og arbeidsgiverenhet i trygde- og skattevesenet.

Bedrifts- og foretaksregisteret omfatter i dag (1980) ca. 165 000 foretak (en- og flerbedriftsforetak) og ca. 194 000 bedrifter i alle næringer, unntatt lønnet husarbeid og jordbruk, skogbruk, fiske og fangst uten leid hjelp. I tillegg kommer ca. 100 000 opphørte enheter som beholdes i registeret, men er merket på en spesiell måte (i registerkoden for "tilstand").

1) Kfr. Ot.prp. nr. 2 (1977-78) og Innst. O. nr 47 (1977-78). 2) Bedrifts- og foretaksregisteret er beskrevet i et Arbeidsnotat (IO 79/1) av Tor Nygaard: "Det sentrale bedrifts- og foretaksregisteret. Innhold og muligheter".

I 1976 ble bedrifts- og foretaksregisteret lagt om til direkte aksess (on-line). Bakgrunnen for omleggingen var at omfanget av registeret økte, nye arbeidsoppgaver ble pålagt Registerkontoret, og antall brukere økte. Dette førte til at det stadig ble stilt krav om hyppigere og bedre ajourhold. Periodisk maskinell oppdatering parallelt med løpende manuell ajourføring på ett sett av alfabetiske datalister var arbeidskrevende, slik at kravet til aktualitet ikke kunne oppfylles. Etter omleggingen til one-line er arbeidsrutinene blitt endret. Alle endringer/korreksjoner av bedrifter, kjennemerker og innføring av nye bedrifter behandles nå fra terminalene.

Tidligere stoppet registerarbeidet ved revisjonen, og resten av arbeidet foregikk ved punch og maskin. Nå deltar den enkelte revisor i arbeidet helt fram.

Den viktigste gevinsten ved omleggingen er at registeret til enhver tid er à jour, hvilket gjør kvaliteten bedre. Arbeidsrutinene er blitt mer effektive, og behovet for arbeidskraft og arkivplass er redusert.

Følgende kjennemerker holdes i dag løpende à jour:

- bedriftsnummer (I.NR.)
- foretaksnummer (F.NR.)
- navn
- karakteristikk (avdelingsbetegnelse)
- mellomadresse
- poststed
- kommune
- næring
- registerenhetstype
- tilstand
- eierform
- momsnummer
- forsendelsesadresse
- postnummer/forsendelsessted
- institusjonell sektorkode

Dessuten har en årlige oppgaver for

- sysselsetting
- omsetning

I tillegg kommer en del kjennemerker av spesiell interesse for enkelte fagkontorer.

Momsnummeret er innført i registeret i samband med at opplysningene i avgiftsmanntallet skal nyttes i statistikkproduksjonen og i ajourholdet av bedrifts- og foretaksregisteret.

Bedrifts- og foretaksnummeret har ikke fått den samme utbredelse utenfor Byrået som fødselsnummeret. Dette skyldes dels at behovet for et felles identifikasjonsnummer for økonomiske enheter innen den offentlige administrasjon ikke er så påtrengende som for personer, og dels at enhetsdefinisjonene som regel må avvike. Administrative enheter må ofte defineres på grunnlag av juridiske regler, og slike definisjoner kan være lite anvendelige for statistiske formål. Byrået stoppet i 1970 utlevering av visse opplysninger i bedrifts- og foretaksregisteret om den enkelte bedrift til brukere utenfor Byrået. Det var opplysninger som var innsamlet i henhold til statistikkloven, og som må karakteriseres som statistikk, som ble rammet av utleveringsstoppen. Enkelte av registerdataene, f.eks. identifikasjonsnummeret og foretaksnummeret kan ikke karakteriseres som statistiske individualdata, og disse blir derfor fritt utlevert (til banker, forsikringsselskaper etc.).

Ved etableringen av avgiftsmanntallet for momsen ble det overveid å nytte foretaksnummer som manntallsnummer, men foretaksnummersystemet tilfredsstilte ikke de krav avgiftsmyndighetene stilte.

Foretaksnummeret er imidlertid brukt som identifikasjonsnummer i banker. Dette danner grunnlag for næringsgruppering og sektorinndelingen i Byråets produksjon av den nye bankstatistikken.

I tillegg til det sentrale bedrifts- og foretaksregister eksisterer det enkelte delregistre over bedrifter ved enkelte av fagkontorene i Byrået. Det arbeides nå med å samordne disse registrene. En slik samordning vil bl.a. gjøre det mulig å hindre at enkelte bedrifter blir spesielt utsatt for mange utvalgsundersøkelser.

3. Andre registre

GAB-registeret

I en stortingsmelding (nr. 25 for 1977-78) er det foreslått opprettet tre registre:

- Grunneiendomsregister
- Adresseregister
- Bygningsregister

Disse registrene omtales ofte under betegnelsen GAB-registrene. GAB-systemet skal altså inneholde ett register over grunneiendommer og ett register over bygninger. Dessuten er adresser og andre såkalte stedfestningsdata knyttet til grunneiendommer og bygninger samlet i et eget EDB-register, kalt Adresseregisteret. Dette adresseregisteret vil bl.a. bli tilført den faste grunnkretsinnndelingen og kode for spredt og tett bebyggelse som skal brukes i Folke- og boligtellingsen 1980.

Bygningsregisteret er tenkt etablert ut fra en datainnsamling som ventelig skal påbegynnes i forbindelse med en landsomfattende taksering av bolig- og fritidseiendommer, et skattetakstregister. Dette skattetakstregisteret vil i første omgang dekke bebodde bygninger og er også tenkt som datakilde for et boligregister.

Arbeidsgiver-/arbeidstakerregisteret

Det er etablert et register over arbeidsgivere og arbeidstakere fra 1. mai 1978. Registeret skal drives som et grunnlagssystem, dvs. et system som inneholder felles datagrunnlag for en rekke forskjellige administrative oppgaver, og som skal nyttes av flere forskjellige myndigheter. Registeret skal kunne brukes til forskjellige formål, f.eks. ved

- utarbeidelse av arbeidstaker- eller sysselsettingsstatistikk
- kontroll av arbeidsforhold når sykepenges skal utbetales
- geografisk differensiering av arbeidsgiveravgiften
- preutfylling av lønns- og trekkoppgaver i den framtidige lønsskatteordningen
- preutfylling av oppgjørslister for arbeidsgivere til kommunekasserere/kemnere

Videre kan arbeidsgiver/arbeidstakerregisteret nyttes til kontroll av utbetaling av stønad til arbeidsledige, og sosialkontorene kan bruke registeret ved utbetaling av sosialstønad. Ved hjelp av innrapporterte arbeidsforhold kan disse kontorene lett finne fram til en klients arbeidsgiver og derfra få opplysninger om inntekt.

Bidragfagden kan nytte registeret til å rapportere om bidrag som eventuelt skal trekkes i en arbeidstakers lønn.

Registeret skal inneholde opplysninger om alle arbeidsgivere og arbeidsforhold. Arbeidsgiverne er identifisert ved et ellevesifret nummer og arbeidstakerne ved sitt fødselsnummer.

Hvis arbeidsgiverne ikke melder fra om endringer (inntak i tjeneste eller opphør av tjenesteforhold), vil de kunne ilegges dagbøter.

5. ARKIVSTATISTISKE SYSTEMER¹⁾

A. Arkivstatistikkens grunnidé

Grunnidéen i det arkivstatistiske system er at opplysninger fra individuelle oppgavegivere (eller m.a.o. data eller kjennemerkeverdier for enheter) arkiveres gjennom lengre tid, slik at de er tilgjengelig og kan hentes fram fra dataarkivet etter behov.

Arkiveringsmetoden vil være avhengig av hvor mye data som skal arkiveres. Er det mye data som f.eks. via en telling hentes inn på skjemaer, vil opplysningene om de enkelte enhetene for alle praktiske formål være utilgjengelig i det øyeblikk skjemaene er pakket vekk, fordi "søking" i store skjemabunker er en kostbar sak. Hentes opplysningene løpende inn, f.eks. slik som i den årlige industristatistikken, må skjemaene pakkes vekk etter hvert. Er det derimot lite data og/eller data som hentes inn svært sjelden, kan arkiveringsmetoden være en annen, f.eks. arkivskap etc.

Når det gjelder store datamengder, eller datamasser som det ofte kalles, har datamaskinene gitt nye tekniske muligheter for arkivering som gir muligheter for å ordne, lagre og gjennom søke store datamasser på kort tid og med små kostnader.

Arkivstatistikkens grunnidé var altså bl.a. at kjennemerkeverdier for enheter skulle kunne sammenliknes over tiden. Det er da nødvendig at enhetene er identifiserbare og at kjennemerkene er definert likt på ethvert tidspunkt.

Foruten å sammenlikne kjennemerkeverdier over tiden, kan vi også sammenlikne "på tvers", dvs. på et bestemt tidspunkt: Vi sammenlikner da data hentet inn fra forskjellige kilder om de samme enhetene, f.eks. data hentet inn om personer ved folketelling og ved helseundersøkelser.

Når det gjelder arkivstatistikken, vil enhver ny opplysning (dvs. enhver ny kjennemerkeverdi eller tall) som blir arkivert, ikke bare ha verdi (informasjonsverdi) i seg selv, men den vil samtidig øke verdien til de andre opplysningene som allerede er arkivert og som den nye opplysningen kan kobles sammen med. Vi kan tenke i et bilde: En motor kan ha stor verdi (i kroner f.eks.) i seg selv, og en båt likeså, men hvis motoren blir plassert i båten, vil båten verdi øke, øke med mer enn motorens egenverdi (såfremt båten allerede er bygget for motorer). Hva er en motorbåt uten motor?

Når det gjelder sammenkobling av data, vil dette i større utstrekning bare være mulig hvis vi har data for hele massen - altså totaltelling. Utvalgsundersøkelser vil gi begrensede muligheter for å koble sammen data fra ulike kilder eller ulike tidspunkter fordi det utvalgte antall enheter vil kunne variere fra undersøkelse til undersøkelse, og det felles antall enheter vil da kunne bli heller lite. Likevel er det også eksempler hvor en tar sikte på å koble sammen data innhentet på forskjellige tidspunkter fra et utvalg, fordi hele eller deler av utvalget er holdt uforandret (forbruksundersøkelser, arbeidskraftundersøkelser).

B. Behovet for arkivstatistikk

Den type informasjon det er behov for, bør danne grunnlaget for oppbyggingen av arkivstatistiske systemer. I motsetning til et "gammeldags" statistikkssystem tillater det arkivstatistiske system bearbeidinger som kan tilfredsstille individuelle og spesifikke brukerbehov.

De publikasjoner som Byrået til nå har utgitt har vært laget ut fra det formål å tilfredsstille flest mulige brukerbehov. Spesielle bearbeidinger er det imidlertid flere brukere som vil ha, så vel offentlige organer og institusjoner som private bedrifter og organisasjoner. Slike spesielle bearbeidingsbehov har brukere som i sin virksomhet trenger kunnskap om enheter som ikke er alminnelig etterspurt, og som derfor ikke kommer med i Byråets publikasjoner. Det kan også være nye problemstillinger som ønskes belyst, og som det av den grunn bare er forskere som er interessert i, dvs. en spesialbearbeiding.

1) Dette kapitlet er mer eller mindre ordrett gjengivelse av avsnitt i Artikler nr. 34, skrevet av O. Aukrust og S. Nordbotten: Dataregistrering, dataarkiver og samfunnsforskning. Oslo 1969.

De spesielle bearbeidingene som etterspørres, varierer sterkt fra bruker til bruker, fra enkle tabeller til tallfesting av strukturene i omfattende matematiske modeller. Ofte må dataene som skal nyttes i slike spesielle bearbeidinger hentes fra forskjellige statistikker og/eller på forskjellige tidspunkter. Den mest hensiktsmessige måte Byrået kan tilfredsstille slike spesielle behov på, er å etablere et arkivstatistisk system hvor det er overkommelig å gå tilbake til grunnmaterialet, koble sammen data (for de samme enheter på ulike tidspunkter og/eller fra ulike kilder, statistiske grunnmaterialer), og deretter gjennomføre spesielle bearbeidinger når behovene melder seg.

Det arkivstatistiske system tillater også omgrupperinger av det foreliggende grunnmateriale. Nå er f.eks. pensjonsalderen 67 år, tidligere var den 70 år, og det er nå mer interessant å se på hvordan aldersgruppen over 67 år vil "utvikle seg" i årene framover. For å kunne gjøre det, må en ha tilbakegående tall. Den någjeldende aldersgrupperingen var ikke aktuell tidligere, og derfor har heller ikke publiserte tabeller og annet grunnlagsmateriale blitt gruppert slik at det uten videre kan tilfredsstille brukerbehovet slik det nå er.

Det er særlig behovet for regional statistikk på spesielle områder som vil kreve at Byrået innfører og vedlikeholder et system som tillater en geografisk klassifikasjon av statistiske enheter ut over det som Standard for handelsområder gjør; både mer detaljert og annerledes, f.eks. i form av et finmasket koordinatnett (landet deles opp i ruter, koordinater, som kan være svært små, f.eks. helt ned til det enkelte hus). Et slikt utviklingsarbeid er nå i gang (Geodatasystemet).

C. Arkivsystemer - databanker

Hovedhensikten med å lagre de innsamlede data slik vi har beskrevet ovenfor, det arkivstatistiske system, er at vi lett kan finne fram til og trekke ønskede data ut av arkivet. De dataarkiver Byrået søker å etablere er både mer omfattende og til dels forskjellig fra det som går under navnet databanker fordi:

1. Til det store antall enheter i dataarkiver knyttes mange kjennemerker over hele livshistorien, slik at hver enhet kan være beskrevet av tusenvis av tegn eller enkelttall.
2. Byrået, i motsetning til dem som arbeider med databanker, får mye data ordnet i puljer som et biprodukt av administrative prosesser.
3. Byrået ønsker ikke å nytte opplysninger om de enkelte enheter, men om hele masser (grupper) av enheter. Statistikk setter ikke søkelyset på den enkelte enhet, men på grupper av enheter.
4. Fordi Byrået ikke behøver å arkivere eller trekke ut data av arkivet umiddelbart når den registrerte begivenhet skjer eller problemstilling er formulert, vil den tekniske organiseringen av dataene kunne være annerledes enn ved databanker hvor en nettopp har disse behovene.

For at Byråets spesielle krav skal kunne tilfredsstilles, må dataarkivet tilfredsstille følgende tre sentrale krav:

- 1) Det må opprettes og vedlikeholdes en kjennemerke katalog for alle enheter som er arkivert. (Foreløpig er det utarbeidd en slik katalog på Underavdelingen for intervjuundersøkelser. Denne katalogen er like mye en oversikt over hvilke spørsmål som er stilt intervjuobjekter)
- 2) Arkivet må splittes opp i delarkiver, slik at de enkelte delene kan kobles sammen etter behov, men hvor de enkelte delene hører til sine spesielle statistikkområder, så kostnader i forbindelse med arkivering og uttak blir lavest mulig
- 3) Det må opprettes og vedlikeholdes et kartotek over hvordan de enkelte kjennemerker med tilhørende data er arkivert rent teknisk

D. Utnytting av dataarkiver

Dataarkivene har gitt store og nye muligheter for å tilfredsstille informasjonsbehov som tidligere ikke har vært mulig å tilfredsstille. Disse arkivene vil sette Byrået i stand til å utnytte de dataene som samles inn mye bedre enn tidligere.

- 1) Arkivene setter Byrået i stand til å lage tabeller eller andre statistiske bearbeidinger på grunnlag av kryssklassifisering av enheter etter kjennemerker. Det er på denne måten mulig å knytte sammen enheter og kjennemerker selv om opplysningene har vært hentet inn fra forskjellige kilder. En kan ved hjelp av arkivstatistikk la behovene, de interessante problemstillingene, styre hva slags tabeller etc. som bør beregnes. Hvorvidt alle disse tabellene, figurene osv. bør publiseres, er en annen sak, men de nye mulighetene Byrået har fått til kryssklassifisering har nok ført til flere sider i publikasjonene, og det er også laget flere nye publikasjoner, i første rekke mer analysepregede publikasjoner (Statistiske analyser).
- 2) Arkivene setter Byrået i stand til å publisere løpende oversikter når dataarkivene holdes à jour. En kan m.a.o. f.eks. holde folketelling hvert år bare med å "kjøre gjennom" et arkiv. Dette alternativet har vært vurdert i forbindelse med planene for Folketellingen i 1980, men foreløpig er ikke dataarkivene gode nok
- 3) Hele masser av statistisk informasjon kan overføres lettere enn tidligere til brukere, men her må man til enhver tid først og fremst ta hensyn til lovverket
- 4) Det er lettere å analysere interessante problemstillinger når vi har dataarkiv. Tabeller er langt ifra den eneste måten å få oversikt over egenskaper ved enhetene i en masse på; andre måter, som f.eks. regresjonsanalyse (statistisk analyseteknikk), er vel så interessante ut ifra analyseformål

Det er nok i mindre grad enn tidligere bearbeidingstekniske problemer som dominerer analyser av samfunnsvitenskapelig karakter. Imidlertid er det nok riktig å si at den arkivstatistiske idé ennå ikke har slått fullt igjennom, spesielt gjelder dette hvor en opererer med andre enheter enn personer på grunn av problemer med å komme fram til såkalte identiske enheter. F.eks. vil bedrifter opphøre, skifte eierforhold, legge om produksjonen, brenne, fusjonere osv., og det trenges da regler for når en bedrift ikke lenger kan sies å være den opprinnelige bedrift.

II. DE ENKELTE LEDDENE I EN STATISTISK UNDERSØKELSE

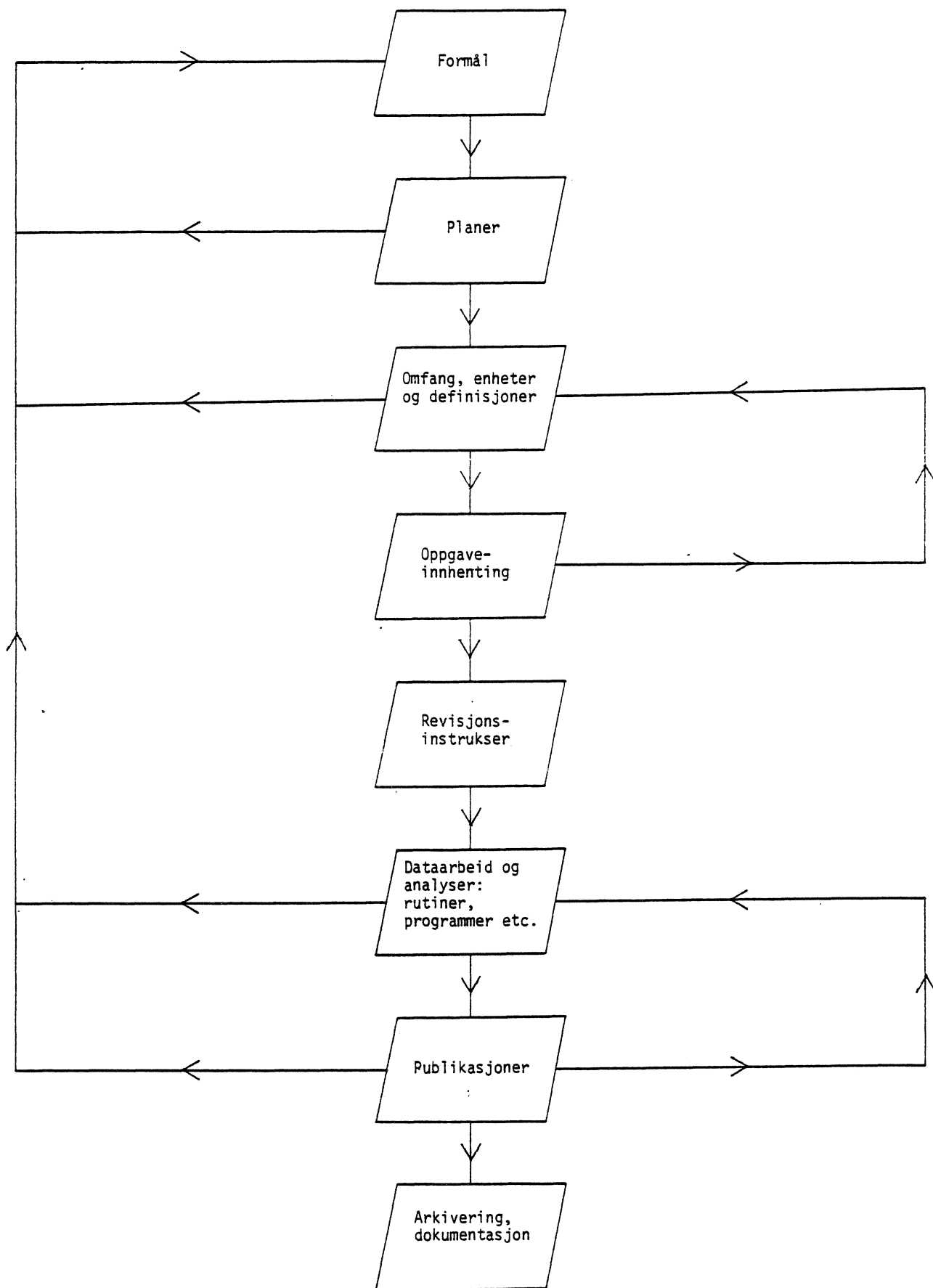
1. SPESIFISERING AV FORMÅLET MED UNDERSØKELSEN

A. Bakgrunnsstudier - hvordan skaffe observasjonsmaterialet?

Kort fortalt kan vi si at statistikk dreier seg om:

- a) Definisjon av enheter og kjennemerker
- b) Avgrensing av masse (undersøkellesbestand)
- c) Registrering av egenskaper (kjennemerker, variabler)
- d) Gruppering av enheter eller kjennemerkeverdier
- e) Opptelling av enheter i grupper (klasser)
- f) Analyse av empiriske fordelinger etter kjennemerker av enheter i massen

For å kunne arbeide med statistiske problemstillinger er det ofte nødvendig å dele opp statistikkproblemet, altså ikke behandle alt under ett. En måte å dele opp på, kan være etter rekkefølgen de enkelte delene vanligvis opptrer i. I figur 1 er en statistisk undersøkelse delt opp i enkelte ledd. Av og til er det ikke nødvendig å gjennomføre en undersøkelse for å få svar på problemstillingen som blir stilt av oppdragsgiver. Statistikk kan allerede være utarbeidd. Her vil vi se på det tilfelle at ny statistikk må utarbeides. La oss først se raskt på figuren.



Figuren må forstås på den måten at starter statistikeren med "Formål", vil han når han er ferdig med formålet med undersøkelsen, gå videre til "Planer". Når han er ferdig med det arbeidet som innbefattes i planleggingen, kan han enten gå videre til "Omfang (etc.)" eller tilbake til "Formål". Hvor planleggeren vil/bør gå, vil avhenge av resultatene av planleggingen sett i forhold til formålet.

Seinere skal vi detaljert ta for oss de enkelte leddene som er vist i figuren. Ved planleggingen av en statistisk undersøkelse må statistikeren på forhånd tenke igjennom og vurdere alle delene som en undersøkelse består av. Når vi i det følgende behandler de enkelte leddene hver for seg, betyr ikke dette at hvert ledd kan vurderes isolert for seg uten hensyn til de øvrige leddene. Hele undersøkelsen må sees under ett, og endringer i ett ledd kan godt føre til endringer i et annet ledd. I hvilken rekkefølge statistikeren bør se på de enkelte leddene, er heller ikke opplagt. Hele undersøkelsen bør sees under ett, og statistikeren bør så vidt mulig når han arbeider med et ledd, også tenke på mulige konsekvenser for andre ledd.

I den første delen av en statistisk undersøkelse, formålsspesifisering, skal en nå fram til en klar og presis problemstilling. Problemet må også formuleres slik at det blir "statistisk". Det er mye "interessant å vite". To forhold kommer da inn og setter grenser, og dette bør være klart allerede under problempresisering:

1. Statistikk er telling av enheter gruppert etter kjennemerkeverdier. Vi må altså kunne fastslå kjennemerkeverdien for alle enheter (kvantitative eller kvalitative). Kjennemerker som ikke har den egenskap, kan ikke brukes. Ofte vil den statistiske undersøkelse derfor bare være et supplement til belysning av en problemstilling. Vi må spørre oss selv: Hvilke enheter og masser skal undersøkes med tanke på hvilke egenskaper?
2. Økonomi, som gjelder både hvilke undersøkelser som skal settes i gang og økonomien innen den enkelte undersøkelse. Det koster som regel ekstra å registrere et ekstra kjennemerke.

For å komme fram til en problemstilling en ønsker å belyse ved hjelp av en statistisk undersøkelse, må en ofte drive et omfattende forarbeid som kan bestå i lesning av litteratur omkring problemet og gjennomgåelse av tidligere undersøkelser. Diskusjon med sakkyndige på området og de viktigste brukerne av statistikken, bør det settes av tid til.

Denne delen av problempresiseringen er meget viktig, og statistikeren bør bruke mye tid på å formulere skriftlig hva han ønsker å belyse ved hjelp av undersøkelsen: Hvilke spørsmål han ønsker å få besvart og hvilke brukere som ønsker en slik undersøkelse.

Jo mer detaljert en greier å presisere problemstillingen, jo lettere blir de øvrige leddene i planleggingen og selve undersøkelsen. Allerede under presiseringsfasen bør en skissere opp hvilke tabeller etc. en vil komme fram til. Altså en klar målsetting for hva en vil fram til, og dette kan med fordel gis i form av tabeller.

Hvis en på forhånd har skrevet ned formålet med undersøkelsen, hva en ønsker å oppnå, kan en etter at undersøkelsen er utført konfrontere resultatene med formålet. Har en derfor ikke oppstilt noe formål for undersøkelsen, kan en heller ikke si om den har vært vellykket eller mislykket. Den ideelle referanserammen er formålet i utgangspunktet og ikke tidligere undersøkelser eller tilsvarende utenlandske undersøkelser, fordi det vil være heller sjelden at de samme krav stilles til to undersøkelser. Gjøres det, er det riktigere å si at det er samme undersøkelse foretatt på to forskjellige tidspunkter.

Når en har valgt og presisert problemstilling, eventuelt skilt ut de delproblemer som egner seg for statistisk analyse, blir neste spørsmål: Hvordan skaffe det observasjonsmateriale som trengs for den statistiske analysen? Finnes opplysninger - statistikk - i form av materiale som andre har samlet inn, er vi på mange måter i en heldig stilling, men det er også ulempes: Vi må ta det som er, kan ikke forme undersøkelsen spesielt for vårt formål. Også når vi samler inn et datamateriale, står vi overfor valgproblemer: Skal vi gjøre bruk av etablerte registre - med de identifikasjonene som der finnes og dermed muliggjøre arkivstatistiske sammenlikninger - eller er kanskje registeret dyrt å bruke eller på andre måter vanskelig tilgjengelig? Likeledes står vi overfor valgproblemet når det gjelder bruk av statistiske standarder hvis de ikke helt klassifiserer enhetene vi er interessert i på den måten vi skulle ønske. Statistikeren bør være meget forsiktig med å vri på standarder eller la være å bruke dem der hvor de finnes. Selv om problemstilling og krav til hva en undersøkelse skal vise endrer seg over tiden, er det grunn til å holde på standardene. Standardene vil som regel gi en basis, og denne basisen kan brukes som grunnlag for ulike analytiske grupperinger. Hva som er analytisk interessant, vil endre seg over tiden.

B. OPERASJONELL UTFORMING AV FORMÅLET, SYSTEMERING¹⁾

1. Hva er systemering?

Med systemering forstår vi arbeidet med utvikling av informasjonssystemer. Begrepet systemarbeid derimot omfatter både arbeidet med informasjonssystemer og andre typer av systemer.

Med et informasjonssystem forstår vi et system som består av informasjonsmengder og informasjonsprosesser. Informasjonsmengdene får vi ved å stille sammen data, statistiske data (tall) og annen mer kvalitativ informasjon. Som vi husker fra tidligere, sier vi at dataene vi samler inn representerer informasjon. Informasjonsmengdene bruker vi så når vi skal løse problemer eller stå i beredskap for å løse mulige problemer.

Et informasjonssystem vil generelt inneholde delsystemer (undersystemer) for

- a) innsamling
- b) lagring
- c) bearbeiding og
- d) distribuering

av informasjonsmengder, dvs. de fire siste "boksene" i figur 1. Informasjonsmengdene, eller dataene, anvendes for problemløsninger i et større system, f.eks. et administrativt styringssystem, nasjonalbudsjetteringssystemet eller liknende. Informasjonssystemet inngår altså i et større system, fordi informasjon (data) er man ikke interessert i for informasjonens skyld, men for å bruke den til noe. Brukerbehovet (den første "boksen" i figur 1) må derfor være klart når man skal utvikle et system for behandling av de informasjonsmengder man vil ha.

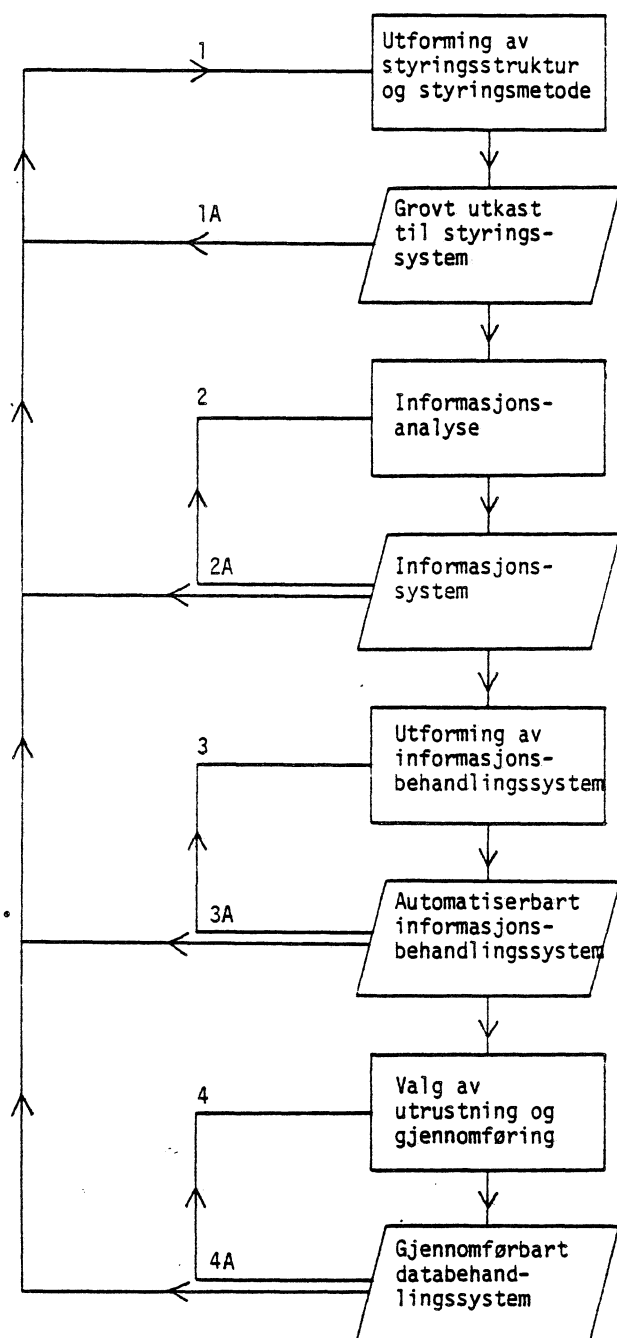
Arbeidet med utvikling av informasjonssystemer eller systemering deles ofte i fire analyse- og metodeområder:

- 1) Utforming av styringsstruktur og styringsmetode
- 2) Informasjonsanalyse
- 3) Utforming av informasjonsbehandlingssystem
- 4) Valg av utrustning (maskiner etc.) og gjennomføring av systemet

I figur 2 på neste side har vi tegnet hvordan de enkelte delene henger sammen.

1) Dette kapitlet er utarbeidd hovedsakelig på grunnlag av et notat av Mats Lundeberg: Informationsanalys. IB-ADB 70 Nr. 34. Denne tegneteknikken (figurer, nummereringer mv.) som Lundeberg har nyttet, er også brukt her, men det vil gå for langt å redegjøre særlig detaljert for den her. Tegningene tar sikte på å få fram sammenhengen mellom informasjonsmengder (romber) og informasjonsprosesser (rektangler). Hver figur skal maksimalt inneholde 10 mengder og/eller prosesser for oversiktens skyld. Er det behov for flere, tegnes dette i form av "forstørrelser" av visse deler av systemet.

Figur 2

Ad 1: Utforming av styringsstruktur og -metode

I dette metodeområdet bestemmer man seg for hvordan man ønsker å styre den virksomhet man skal utvikle et informasjonssystem for. Resultatet blir et grovt utkast til styringssystem. F.eks. ønsker regjeringen å styre landets økonomi. I metodeområde 1 vil vi da utforme det systemet landets økonomi skal styres etter, altså nasjonalbudsjettet.

Ad 2: Informasjonsanalyse

I dette metodeområdet analyserer man de informasjonsstrømmer som det tenkte styringssystemet behøver, f.eks. hvilke data er nødvendig for å utarbeide nasjonalbudsjettet. Man lager en plan for sammenhengen mellom strømmen av informasjonsmengdene. (Vi har en strøm av informasjonsmengder når vi har en statistikk som gjentas med visse mellomrom, f.eks. kvartalsstatistikk.) Informasjonsanalysen gir som resultat en beskrivelse av et informasjonssystem.

Ad 3: Utforming av informasjonsbehandlingssystemet

Her bestemmer man seg for hvordan man skal inn-
samle, lagre, bearbeide og distribuere de informasjonsmengder (data) man skal arbeide med.

Ad 4: Valg av utrustning og gjennomføring

Her velger man helt konkret de hjelpemidler man må ta i bruk for å kunne gjøre det man ønsker. Dessuten utarbeider man i denne fasen bearbeidingsrutiner, dvs. programmer for hvordan informasjonsmengdene skal behandles (innsamles, bearbeides, lagres og distribueres).

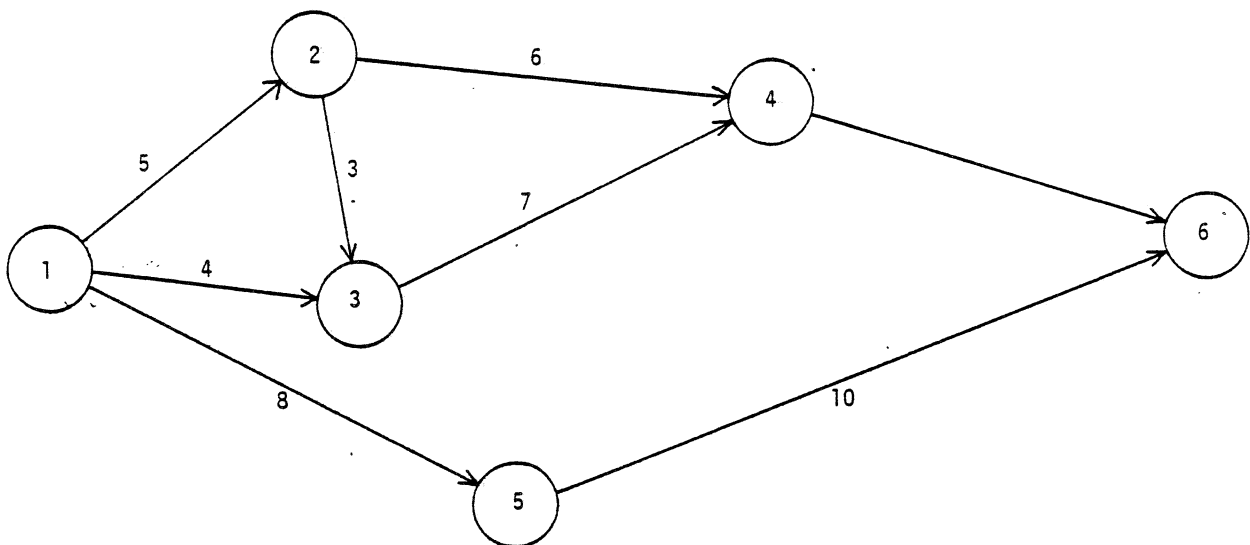
En bestemt statistisk undersøkelse vil da ta utgangspunkt i 2A og 1A, som til sammen gir oss undersøkelsens formål (jfr. figur 1). Statistiske data vil være en del av den informasjon man har bruk for i styringen, og i 3 vil vi da utforme bl.a. hvordan vi skal innsamle data og behandle disse.

En annen teknikk som tar sikte på å være et hjelpemiddel ved forenkling og forbedring av tidsplanlegging av store og omfattende prosjekter med mange operasjoner som til dels er avhengig av hverandre, er såkalt nettverksteknikk eller -analyse. Denne teknikken ble det eksperimentert noe med i Byrået i begynnelsen av 60-årene¹⁾. Nettverksmetoden er særlig et hjelpemiddel til systematisk og effektiv kontroll og oppdatering av planer under gjennomføringen. Metoden er særlig i bruk ved byggeprosjekter for å sikre at materialer og kvalifisert mannskap (murere, malere osv.) er tilstede på rett tid.

Første steget i en nettverksanalyse er et nettverksdiagram. Nettverket består av en figur hvor de enkelte operasjoner som inngår i prosjektet er forbundet med hverandre. Også her er det laget bestemte regler for tegning av nettverkets punkter (sirkler) og piler. Et punkt betegner en hending, at noe er gjort; en pil betegner en aktivitet, altså en prosess som krever tid, og denne tid er avmerket med et tall langs pilen (f.eks. antall dager). Hver hending gis et nummer, og i figur 3 er det vist et eksempel på et meget enkelt nettverk med 6 hendinger og 8 aktiviteter. Ut fra startpunktet 1 utføres "aktiviteter" som tar henholdsvis 5, 4 og 8 dager, og vi ender i "hending" 2, 3 og 5. Etter "hending" 2 er det to "aktiviteter" som skal utføres og som tar henholdsvis 3 og 6 dager, osv. En får på denne måten en god oversikt over hvilke "aktiviteter" som avhenger av hverandre og hvor det rent tidsmessig er "lønnsomt" å sette inn ressurser. I figur 3 ser vi at det er veien 1, 2, 3, 4, 6 som tar lengst tid (21 dager), mens veien 1, 3, 4, 6 tar 17 dager og 1, 5, 6 tar 18 dager. Uansett hvor mye vi kutter ned på tiden som går med til aktivitetene på veien 1, 5, 6, vil prosjektet kreve 21 dager, og følgelig er det aktiviteter langs veien 1, 2, 3, 4, 6 som i første rekke bør vurderes nærmere.

I tiden som de enkelte aktivitetene antas å bruke, kan det også ligge ventetid og annen slark. Den lengste veien kalles ofte for den kritiske veien.

Figur 3



1) Se Arbeidsnotat IO 65/1: Nettverksanalyse i statistikkproduksjonen av N. Bakke, E. Homb og S. Nordbotten.

2. Hvordan systematisere datainnsamlingen?

Hvis de data vi samler inn skal kunne innpasses og tilfredsstille flere ulike styringsbehov, f.eks. behov som det offentlige har og behov som bedriftene har, vil dette ha stor betydning for datainnsamlingen. For å kunne tilfredsstille flere ulike behov, vil det være nødvendig å samle inn dataene slik at de tilfredsstiller kravene til et arkivstatistisk system. Nettopp når vi samler inn data innenfor en slik systematikk, vil vi ha muligheter for å kombinere informasjonsmengder, idet vi da har - som vi husker - identifisert enhetene og dataene.

Det er videre klart at vet man ikke hvilket (styrings-)problem som skal løses, vil det kunne være bortkastet å samle inn data. En bør vite på forhånd hva man skal med dataene, hvilket system de skal representere informasjon i. Dette systemet kan selvfølgelig være et antatt framtidig problem.

Systematisering av datainnsamlingen gjennomføres best ved bruk av eksisterende standarder, bruk av veldefinerte begreper (enheter og kjennemerker) og hvis mulig, ved å ta utgangspunkt i etablerte og ajourholdte registre.

2. PLANLEGGING AV GJENNOMFØRINGEN AV UNDERSØKELSEN

A. Arbeidsplan, opplegg av undersøkelsen

Som nevnt innledningsvis må statistikeren som planlegger en undersøkelse, se de enkelte leddene i undersøkelsen i sammenheng, men hans planleggingsarbeid må selvfølgelig starte et sted. Han bør derfor enten begynne med de(t) ledd som har flest konsekvenser for de øvrige, deretter de(t) som har nest mest osv. På den måten vil han arbeide seg gjennom de enkelte leddene.

Alternativt kan statistikeren begynne det første ledd, "første" sett i logisk sammenheng eller tidssammenheng, og så arbeide seg gjennom alle leddene en gang. Når han er ferdig med det, kan han begynne forfra igjen, idet han nå "vet" noe mer enn da han startet planleggingen. Statistikeren kan da gå gjennom alle delene en gang til, en iterasjon til, som denne prosedyren kan kalles. Hvor mange iterasjoner, dvs. hvor mange ganger han behøver å gå gjennom de enkelte leddene, vil avhenge av selve undersøkelsen, hvor komplisert og omfattende den er. I noen av iterasjonene er det ikke sikkert at statistikeren bør gå gjennom alle leddene før han begynner på nytt igjen. Det kan hende at de resultater han kom fram til mht. siste ledd vil medføre store forandringer for de(t) første ledd, slik at det ikke har noen hensikt å gå videre på de neste ledd før de(t) første er endret.

I figur 1, hvor de enkelte leddene i en statistisk undersøkelse er tegnet inn, er planleggingen plassert tidlig. Planleggingen i seg selv innebærer at statistikken går gjennom samtlige ledd "på papiret". For mest mulig realistisk å gå gjennom de enkelte ledd på forhånd, altså "på papiret" før undersøkelsen utføres, kan det gjennomføres en prøveundersøkelse. En prøveundersøkelse er ofte en fin hjelp når en skal tilrettelegge undersøkelsen og gå gjennom alle leddene på forhånd på papiret. Planleggingsarbeidet skal i prinsippet føres helt fram til oppstilling av de endelige tabeller og utregninger av de statistiske mål. Men ofte vil vi måtte endre opplegget når vi kjenner deler av materialet, f.eks. gjennom en prøveundersøkelse. Vi må under planleggingen spesielt huske at detaljer kan aggregeres (summeres etc.), men aggregater kan ikke deles, og at detaljer koster. Ber vi f.eks. om bedriftens samlede utbetalte lønn, kan vi ikke på det grunnlag lage statistikk over lønn til arbeidere og funksjonærer hver for seg (uten spesielle beregninger som må bygge på forutsetninger), men ber en bedriftene om funksjonær- og arbeiderlønn hver for seg, kan en lett også utarbeide oversikter over samlet lønn.

Planene må omfatte mønsteret for en undersøkelse. Har en dette mønsteret, har en også en arbeidsplan, men dessuten må det også stilles opp særskilte tidsplaner og kostnadsplaner.

For å komme fram til mønsteret for en undersøkelse, må vi se nærmere på hvilke muligheter vi har for å finne fram til enhetene i den massen vi skal undersøke og å registrere dens egenskaper. F.eks. skal vi lage en undersøkelse som viser fordelingen av norske husholdningers forbruksutgifter. Hvordan lar så dette seg gjøre? La oss kort se på noen problemstillinger i denne forbindelse:

Må vi registrere forbruket hos alle enhetene i massen, eller kan vi finne et representativt utvalg? Dette vil igjen være et spørsmål om vi har register som gjør det mulig å få tak i alle enhetene eller trekke et utvalg. Skal vi be om opplysninger gjennom intervju, eller skal vi be husholdningene føre regnskap? Hvis vi bruker intervjuere, må de møte opp og spørre. Er det da noen hjemme, husker de det intervjuerne spør om, kan de og/eller vil de svare f.eks. om brennevin og tobakk? Hvis vi bruker regnskap, må det bety at vi tror husholdningene vil føre regnskap lenge nok til at vi får et riktig bilde. Er det f.eks. bare de "bra" husholdningene som kan og vil føre regnskap? Påvirker regnskapsføringen forbruket? - Uansett om vi velger at husholdningene skal gi svar til en intervjuer eller selv sende inn et regnskap, må vi ta stilling til hvor lang periode vi ønsker opplysninger for: En dag, en uke, en måned, ett år osv. I de løpende forbruksundersøkelsene som nå er i gang, er det en kombinasjon av intervju (innlednings- og avslutningsintervju) og regnskap (for en 14-dagersperiode). De fleste husholdningene som blir trukket ut (et årlig utvalg på 1 500 som fordeles over hele året), har ikke vært med før, bare noen husholdninger var med året i forveien.

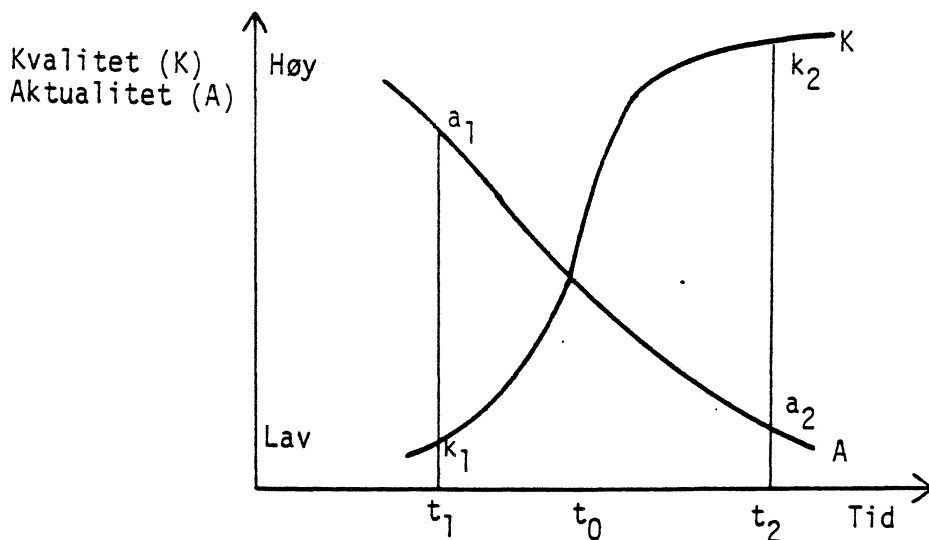
Planleggingsarbeidet vil stille seg forskjellig etter som det gjelder a) engangs-, b) sjeldnere eller c) regulære (særlig årlige og kvartalsvise) undersøkelser. Ved de regulære, spesielt månedlige og kvartalsvise undersøkelser, gjelder det å bygge opp faste rutiner. Ved engangs- og sjeldnere undersøkelser bygger en på mønsteret i tilsvarende undersøkelser.

Både når man bygger opp faste rutiner eller gjør bruk av opplegg for andre undersøkelser, er det en fare for at vi blir hengende fast i foreldede opplegg som bl.a. ikke tar hensyn til endringer i samfunnsforhold og teknisk utstyr.

B. Tidsplan

Under planleggingen er det viktig i første rekke å få fastlagt krav som stilles til kvalitet og aktualitet. En enkel figur illustrerer dette forholdet:

Figur 4



En undersøkelse som "gir" resultater etter t_1 dager (eller måneder), har lav kvalitet (k_1), men høy aktualitet (a_1). Lar vi statistikeren arbeide mer med undersøkelsen, f.eks. revidere grundigere, purre flere ganger osv., vil tiden gå, og etter t_2 dager vil aktualiteten ha sunket (a_2), men datakvaliteten ha økt (k_2). Selv om vi har holdt på med en undersøkelse i en lengre periode, vil det nesten alltid være mulig å øke kvaliteten ved ytterligere kontroller og bearbeidinger. Imidlertid blir det som regel stadig dyrere - eventuelt vil det ta lengre tid - å øke kvaliteten ut over et visst nivå. Formen på K-kurven antyder dette og antyder også at bearbeiding lengre enn til to ikke er å anbefale.

En tidsplan skal avklare når de forskjellige ledd skal utføres og hvordan de skal passes inn i hverandre. Nettverksteknikk er meget anvendbart for slik planlegging. Tidsplanleggingen må ikke bare gi planer for når statistikeren skal gjøre sine forskjellige arbeidsoppgaver, men også plassere undersøkelsen i tid for oppgavegiverne og brukerne: Når passer det best at vi spør bedriftene om regnskapene for siste år, når må lønnsstatistikk være utarbeidd for at den skal kunne brukes i lønnsforhandlinger, osv.? Den type statistikk vi får i oppdrag å skaffe, vil klart stille krav til tidsplanen: Tall til belysning av konjunkturutviklingen må være aktuelle, mens tall til belysning av strukturelle forhold (industristruktur), som skal forklare visse hendinger (hvem stemte på hva i EF-avstemningen?), haster ikke så mye.

Det kreves ganske grundig planlegging for å få gjort rett operasjon av rette vedkommende på rett tidspunkt; f.eks. at revisjon av datamaterialet påbegynnes og avsluttes på bestemte datoer som er tilpasset undersøkelsen for øvrig, at revisjonen utføres av kvalifiserte revisorer, etc.

C. Kostnadsplan

Foruten å planlegge hvordan undersøkelsen skal gjennomføres, dvs. gå inn på de enkelte ledd i figur 1, må statistikeren i tillegg til en tidsplan sette opp kostnadsplaner. I første rekke vil det være et finansielt budsjett, men det har blitt, ikke minst i Byrået, stadig viktigere også å stille opp arbeidskraftbudsjett og maskinbudsjett.

Ofte avsettes det et bestemt beløp til en undersøkelse. Planleggeren må fordele dette beløpet på de enkelte delene av undersøkelsen, slik at sluttresultatet blir best mulig. Her må alle leddene sees i sammenheng med hverandre. En får lite hjelp av å bruke mange penger på å konstruere fine skjemaer med mange interessante spørsmål, hvis bare noen blir bearbeidd og/eller bearbeidingen er overfladisk. Det er lite lønnsomt å finrevidere skjemaer hvis det endelige resultatet kan bli bedre hvis noe av pengene som blir brukt til revisjon, i stedet blir brukt til å forbedre skjemaene, osv.

Hvor detaljert kostnads- og tidsplanene bør være, vil avhenge mye av hvilken type undersøkelse som skal utføres. Først bør det utarbeides foreløpige planer som bør revideres etter en "gjennomløping" av alle leddene i undersøkelsen, deretter en mer detaljert og muligens enda en helt til slutt. Under gjennomføringen av selve undersøkelsen vil det trolig bli behov for revisjon av de opprinnelige anslagene, men dette bør ikke bety at kostnads- og tidsplanleggingen ofres lite tid til å begynne med.

Arbeidskraftbudsjettet bør angi hvilke typer av arbeidskraft (assistenter, saksbehandlere, akademikere osv.) som forutsettes brukt og når de skal være i arbeid. Er det nødvendig å lære opp personale? Vil intervjustaben være tilgjengelig?

Maskinbudsjettet skal angi hvilke maskiner og bearbeidingsopplegg som forutsettes nyttet. Skal datamaterialet bearbeides manuelt ved hjelp av bordregnemaskiner eller ved hjelp av programmerbare elektroniske regnemaskiner?

Også når det gjelder kostnadsplanlegging, vil en prøveundersøkelse gi verdifullt erfaringsmateriale. Jo større og mer omfattende undersøkelsen er, dess mer kreves det av planlegging. Statistikk er vanligvis masseoperasjoner. Hver enkelt operasjon i en undersøkelse er fra ethvert synspunkt liten; liten i tid, omfang og penger, men det er så mange av dem at det kan bli store summer. Folketellingen 1970 kostet om lag 40 millioner kroner, det vil si om lag 10 kroner pr. person, og en økning i kostnadene pr. person på 1 krone gir hele 4 millioner kroner i økte kostnader. 1 sekund ekstra bearbeiding pr. person blir over 1 000 timer (1/2 årsverk). Det blir altså meget kostbart å gjøre feil. Derfor er det viktig at planen er god og dekker alle mulige enkeltheter.

3. OMFANG, ENHETER OG DEFINISJONER

A. Identifisering av masse og enheter

I denne delen av en statistisk undersøkelse må vi fastlegge hvem som skal være med, dvs. hva undersøkelsen skal omfatte. Vi må avgrense den massen vi skal undersøke, slik at vi vet hvilke enheter som skal med og definere disse enhetene hvis det ikke er helt opplagt hva enhetene står for. Statistiske enheter (klassifikasjonsenheter) kan være så mangt, f.eks. terningkast, trekninger av tall, mennesker i Norge, fødsler i Norge, passasjerer på jernbanen, etc.

Det er her viktig å være oppmerksom på at vi i en og samme undersøkelse kan operere med flere sett av enheter. Eksempelvis kan man i en og samme telling operere med fire sett enheter: individer, familier, husvære (leiligheter) og hus. I en bedriftstelling kan man operere med både foretak og bedrift som enhet. Ved folketellingene blir det samlet inn opplysninger om den enkelte person (alder, ekteskapelig status, utdanning mv.), husholdningen som personen tilhørte, den leilighet eller husvære som familien bodde i og om det huset hvor leiligheten befant seg.

I samband med denne delen av undersøkelsen er det en fordel å ha et register, og da helst et som er permanent.

Det er mange problemer som oppstår når vi skal avgrense masse og definere enheter, men problemene er ofte svært spesielle: Hvis vi f.eks. skal ha en statistikk over industrien i Norge, skal vi da ta med verksteder? Liknende eksempler på avgrensinger er: Mennesker som befinner seg i Norge eller på norske skip 31. oktober 1970, osv. Vi kan imidlertid ikke alltid undersøke nøyaktig den masse vi egentlig er interessert i. Vi må konkretisere massen. Eksempel: Vi er interessert i forbruket hos bønder. Hvem er bonde? Hva med fiskere og småbrukere? Vi må finne en operasjonell definisjon av enheten som gjør det mulig å finne fram til alle enhetene som tilhører massen. Her kan det imidlertid lett bli konflikt mellom vår interesse for en masse og våre muligheter for å identifisere og registrere enhetene i massen. Dette kan tvinge oss til å arbeide med en annen masse enn den vi egentlig hadde tenkt. To eksempler belyser dette:

K r i m i n a l i t e t: Vi er interessert i kriminalitet og ønsker å telle opp forbrytelser og forseelser begått av personer med forskjellig alder og andre sosiale kjennemerker. Imidlertid kan vi bare registrere massen av anmeldte forbrytelser, og vi kan bare fordele etter forbrytternes egenskaper for oppklarte og til dels rettslig behandlede forbrytelser. Viser det seg f.eks. at ingen akademiker har begått en oppklart forbrytelse, er det ikke dermed sagt at akademikere ikke begår forbrytelser.

A r b e i d s l ø s h e t: I den offisielle statistikken til Arbeidsdirektoratet registreres bare de som melder seg arbeidsløse på arbeidskontorene. I Byråets Arbeidskraftundersøkelser (AKU) tar en også sikte på å få opplysninger om den "virkelige" arbeidsløsheten, dvs. de som er arbeidsløse, og som har gjort noe for å få arbeid, men som ikke nødvendigvis melder fra til arbeidskontorene (ILO-definisjonen). Av og til opptrer også begrepet "skjult arbeidsløshet", og det står for alle som ikke er yrkesaktive og som ikke har gjort noe for å få arbeid fordi de ikke har håp om å få passende arbeid.

Samtidig med at vi fastlegger masse og enhet (eventuelt enheter), må vi bestemme oss for hvilken periode eller hvilket tidspunkt undersøkelsen skal omfatte. Det gjelder å finne et tidspunkt, eller en periode, som passer både for telleenhetene og som passer inn i formålet. Det kan derfor være nødvendig å gå nøye tilbake til formålet mens vi behandler dette punktet. (Jfr. pilen på figur 1, s. 35.)

B. Spesifikasjon av problemrelevante egenskaper

Når vi har fått klarlagt masse og enheter, må vi klarlegge hva vi vil vite om enhetene og massen. Enhver enhet må lokaliseres i forhold til de egenskapene vi er interessert i. Som vi husker kalles de egenskapene vi vil karakterisere enhetene med for kjennemerker, og en statistisk undersøkelse utføres for å få innhentet verdier på disse kjennemerkene, data.

Hvilke typer av data vi vil samle inn, avhenger helt av det aktuelle problemet vi vil belyse. Kjennemerkeverdien kan enten være en av noen få verdier (ja eller nei; gift, før gift, osv.) eller en av et uendelig sett av verdier (størrelse på inntekt, osv.). Hvilke alternative verdier kjennemerkene kan ha, må spesifiseres på forhånd. F.eks. er det noe annet at 27 prosent av befolkningen bor i eneboliger enn at 41 prosent av husene er eneboliger. Vi må klarlegge på forhånd hvilke egenskaper vi er interessert i.

4. OPPGAVEINNHENTING

A. Metoder for innhenting, valg av tellemåte

Det er fire hovedtyper av tellemåter, undersøkelsesmetoder:

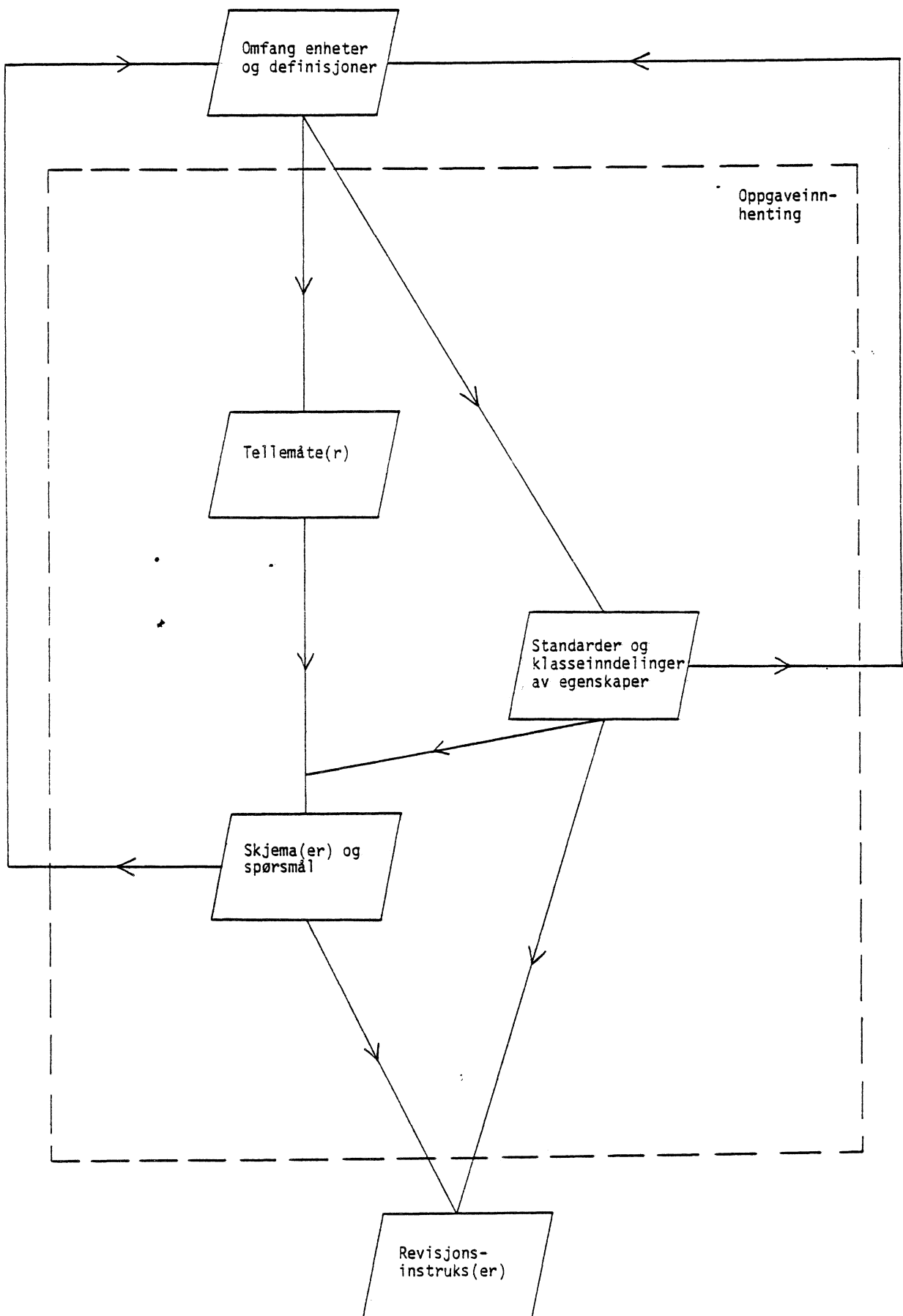
1. Postundersøkelser som er registreringer hos og ved telleenhetene (klassifikasjonsenheten eller registreringsenheten). Telleenheten gir svar på et skjema som sendes i posten
2. Kontakt med telleenheten via andre kommunikasjonsmidler enn post, dvs. telefon, telegraf etc. Dette er en mer direkte kontaktform med telleenheten enn postundersøkelse
3. Intervjuundersøkelser som består i registreringer hos telleenheten ved spesielle tellere, intervjuere
4. Registerundersøkelser, enten fra administrative eller statistiske registre. Ofte vil dette være at det utarbeides statistikk på grunnlag av administrative rutiner som får data fra sine "klienter" på en systematisk måte. Byrået får da dataene fra den administrative enheten som regel via post

Vi skal i det følgende komme nærmere inn på de enkelte tellemåtene. Hvilken måte vi skal nytte i en undersøkelse for å få de ønskede data, vil avhenge av en rekke forhold. Det er dessuten også mulig å nytte flere av de fire metodene for en og samme undersøkelse, f.eks. postundersøkelse primært, mens de som ikke svarer (eller et utvalg av disse) kan bli intervjuet. Mellom de fire hovedformene kan det være alle slags mellomformer og kombinasjoner. Tellemåten vi skal bruke vil først og fremst avhenge av de penger og den tid vi har til disposisjon. Er disse to faktorene gitt for en undersøkelse, bør vi velge den tellemåte som gir det beste ferdige statistikkprodukt. Vi må da huske å se dette ledd i sammenheng med alle de øvrige ledd i undersøkelsen.

Foruten den tid og de penger vi har til disposisjon, vil valg av tellemåte kunne avhenge av hvem som telles (dvs. klassifikasjonsenhetene), hva vi skal registrere hos disse enhetene (dvs. kjennemerkene) og hvem vi får opplysningene fra (dvs. registreringsenhetene). Det kan f.eks. hende at enhetene ikke kan gi opplysningene på "stående fot", at enkelte enheter bør ha bedre rettledning enn andre, at enkelte ikke kan lese norsk, at enhetene ikke vil at andre enn et "upersonlig" kontor skal få vite noe om dem (dvs. enhetene ønsker ikke å "utlevere" seg til intervjuere), osv.

Den tellemåten vi velger, vil kunne ha stor betydning for kvaliteten på undersøkelsen. Svarprosenten vil avhenge av tellemåten, likeså selve svarenes kvalitet. Ofte vil imidlertid en tellemåte som gir høye svarprosent og gode svar, også koste mer. Hvis spørsmålene vi stiller eller måten vi stiller dem på gir irritasjon, vil svarenes godhet eller pålitelighet kunne bli dårlig. Det kan være fristende for Byrået eller andre å skyve kostnadene over på andre, dvs. at man velger den metode som er billigst for oss (som regel postundersøkelse), men vi må da huske på at dette kan øke kostnadene for telleenhetene. Ved valg av tellemetode må vi også ta hensyn til oppgavegivers registreringskostnader og ikke bare våre egne kostnader. Trolig er vi på grunn av våre egne kostnader også mer forsiktig med å stille unødige spørsmål når hvert spørsmål koster så mye som tilfellet er med intervjuundersøkelser, men helt sikkert er det ikke, fordi de faste kostnadene pr. intervju utgjør en stor del av de totale kostnadene ved intervjuundersøkelser.

Figur 5. En forstørrelse av leddet "oppgaveinnhenting" i en statistisk undersøkelse



Av og til hevdes det at stikkprøveundersøkelser er en form for datainnsamling, men det kan synes søkt å kalle dette en egen form, idet slike stikkprøveundersøkelser må foretas ved hjelp av enten post, telefon, intervju eller registre, dvs. ved hjelp av en av de andre metodene. I virkeligheten er stikkprøveundersøkelse en form for utvalgsundersøkelse. En bruker gjerne begrepet stikkprøveundersøkelse på utvalgsundersøkelser innenfor industrien. Ved kvalitetskontroller o.l. av produksjonen tas stikkprøver. Utvalgsundersøkelser som tar sikte på å måle politiske meninger i en befolkning eller befolkningsgrupper, kalles gjerne gallupundersøkelser. Undersøkelsesmetoden her er som regel intervju. Når aviser o.l. ønsker å få rede på folks meninger om et aktuelt spørsmål, stilles spørsmål til et svært lite utvalg ("Fire i forbifarten" o.l.); slike undersøkelser kalles ofte for enqueter eller galluper.

Innsamlingsmetoder vil vi definere som de metoder som nyttes til å få den ønskede informasjon om virkeligheten overført og lagret som data på passende databærere (f.eks. på et skjema, magnetbånd, hullkort osv.). Når vi har fått dataene tilgjengelig på databærere, vil dataene kunne bearbeides videre.

1. Postundersøkelser

Flertallet av Statistisk Sentralbyrås undersøkelser er postundersøkelser.

Det største problemet ved postundersøkelser er: Hvordan skal vi få enhetene til å svare? Vi har ingen direkte kontakt med telleenheten, bortsett fra spørreskjemaet og eventuelle følgeskriv. Enhetene kan enten ikke ønske å svare eller de kan rett og slett ikke svare (f.eks. forstår de ikke spørsmålene). Ved postundersøkelser kommer vi som regel lengst når vi har en lovhjemmel, men vi har også eksempler på frivillige undersøkelser med meget høy svarprosent (f.eks. har både den kvartalsvise investeringsstatistikken og konjunkturbarometeret over 95 prosent i svarprosent). Oppgaver til offentlig statistikk innhentes ofte i medhold av stortingsvedtak og kongelig resolusjon som er gitt med hjemmel i statistikkloven av 25. april 1907. Har vi lovhjemmel, kan vi tvinge enhetene til å svare, men ved eventuell tvang vil kvaliteten på de svarene vi får kunne være dårlig. Hvor langt Byrået f.eks. kan gå med hensyn til å få akkurat de oppgavene det ber om, selv om enhetene må bruke mye tid og penger (arbeidskraft) for å besvare spørsmålene, er et problem som oppstår fra tid til annen, uten at et generelt svar kan gis.

En lovhjemmel pålegger oppgavegiverne å sende inn blanketten i utfylt stand. Unnlattelse av å gi lovhjemlet oppgave er et lovbrudd på linje med andre lovbrudd (f.eks. brudd på trafikkregler). Loven pålegger statistikkprodusenten å nytte opplysningene bare i statistisk øyemed, og opplysningene må ikke offentliggjøres på en slik måte at det kan være skadelig for oppgavegivernes interesser.

Enten man har lovhjemmel eller ikke, vil det være en fordel ved oppgaveinnhenting at den anbefales av organisasjoner som enheten er knyttet til. Ofte vil det også ved postundersøkelser være en fordel at undersøkelsen omtales fordelaktig i massemedier på forhånd (PR, goodwill etc.). Faren ved at det skapes en opinion om undersøkelsen på forhånd er at svarene som enheten gir, kan bli farget av det som skrives om undersøkelsen.

Det er viktig å få oppgavegiverne positivt interessert. Derfor utarbeides det som regel ved postundersøkelsen et følgeskriv, som redegjør nærmere for hensikten med undersøkelsen. Ved en privat undersøkelse vil alt stå og falle med dette følgeskrivet. Det er flere eksempler på undersøkelser som har kommet galt av sted fordi det var gitt dårlig forhåndsinformasjon (et eksempel er en sosiologisk undersøkelse i Oslo som gikk ut på å følge en aldersklasse av gutter i en årrekke).

Lovhjemmel eller ikke, det er av sentral betydning at oppgavegiverne har tillit til innsamlingsorganisasjonen. Skal en få folk, bedrifter etc. til å gi fra seg riktige opplysninger, må ikke opplysningene bli brukt på en slik måte at de kan skade dem selv.

Også ved postundersøkelser gjelder det at enhetene ikke må treffes på et ubeleilig tidspunkt. Utsendelse av skjemaer etc. i ferietiden vil kunne redusere svarprosenten svært mye. Enhetene må også få rimelig tid på seg til å besvare spørsmålene.

Vi må på forhånd være klar over hva vi skal gjøre hvis vi ikke får inn oppgavene fra alle enhetene, samtidig som vi må være klar over hvor stor oppslutning (svarprosent) vi må ha for at vi skal kunne trekke slutninger fra undersøkelsen. En masse med store variasjoner i svarene må ha stor oppslutning; venter vi lite variasjon i svarene, kan vi tolerere litt større frafall. Ønsker vi mange detaljer, f.eks. kommunetall og ikke bare landstall, må vi ha større oppslutning.

Hvis vi har valgt postundersøkelse som innhentingsmetode, har vi svart ja på spørsmålet om enhetene kan besvare spørsmålene i skjemaet uten hjelp (dvs. uten hjelp fra intervjuer). Dessuten har vi da også tatt stilling til hva som skal gjøres hvis vi oppdager at enheten har svart feil. Ved intervjuundersøkelser kan en del feil avverges og/eller opprettes allerede under intervjuet, mens ved en postundersøkelse vil feilene først kunne oppdages under revisjonen.

Av og til sendes ikke skjemaene direkte til enhetene, men til andre offentlige myndigheter, enten for at de skal svare på opplysninger om enhetene (f.eks. skattestatistikk) eller at de sender skjemaene videre til enhetene (f.eks. til folkeregistrene som så sender det videre). Dette kan være den beste metode hvis disse andre offentlige myndighetene har de best ajourførte adresseopplysninger, og dessuten kan Byrået da få opplysningene direkte på bånd fra det administrative organet. Etter hvert som Statistisk Sentralbyrås registre har blitt etablert og ajourholdt, har bruk av andre offentlige etaters adresser avtatt. Folketellingsskjemaene for 1970 fikk f.eks. identifikasjonene om enhetene preutfylt i Byrået, men de lokale folkeregistrene kontrollerte opplysningene mot sitt eget register.

Ved postundersøkelser opplever man som oftest ved fristens utløp, nesten uansett hvor lang tid det er mellom utsending av skjemaene og innsendingsfrist, at bare ca. 50 prosent av enhetene har svart. Derfor må enhetene både påminnes og purres. Ofte sendes det ut en påminnelse rett før innsendingsfrist og to til tre purringer etterpå. Etter hvert som tiden går bruker purringene å bli strengere i teksten, og har man lovhjælp, vil en til slutt true med å gå til politianmeldelse hvis enheten ikke svarer. Ofte foretas en siste purring via telefon før eventuell anmeldelse.

Generelt er en avhengig av god respons på den første henvendelsen. Ved å sende purrebrev og ved å vente til utover svarfristen, vil flere svare. Dette gir økende kostnader og avtakende aktualitet. Normalt er det nødvendig å purre bare på en del av oppgavegiverne, og det gjelder å finne fram til en purreform som sikrer en bedre reaksjon enn den første henvendelsen. Purrebrevene må utformes slik at de blir lagt merke til. Siden skjemaene som regel må sendes ut i god tid før svarfristen, vil det ofte være en fordel med en påminnelse rett før fristen utgår. De fleste oppgavegiverne ønsker å være lovlydige, og en høflig oppfordring har vist å ha god effekt.

Ved årvisse, gjentatte undersøkelser er det en fare for at oppgavegiverne blir så vant til purrebrev o.l. at de ikke egentlig legger merke til det som skrives. Av den grunn bør endringer i rutiner eller purreform skje fra tid til annen.

Postundersøkelser er en særlig anvendelig undersøkelsesform når vi har fortrolige (konfidentielle) og/eller intime spørsmål (f.eks. utdanning, inntekt), når vi har få og enkle spørsmål, når telleenheten er vant til skjemaarbeid (f.eks. bedrifter) og når tellingen inngår i en regelmessig rutine, slik at oppgavegiverne lærer fra gang til gang.

2. Telefon-, telegrafundersøkelser

Av og til vil det være behov for hurtig svar, og da kan telefon, telegraf eller telex brukes. Helst bør også antall telleenheter vi skal ha svar fra, være lite. Er det et lite antall telleenheter som bor spredt, kan telefon av og til med fordel nyttes istedenfor intervjuer pga. kostnadene. En tredje fordel med telefonundersøkelse er at vi her har direkte kontakt med telleenheten.

Bruk av slike hurtige kommunikasjonsmidler krever selvfølgelig at de intervjuede har telefon eller kan nås via telegraf etc.

Aviser og andre nyhetsmedier bruker ofte telefon rundt omkring til sine avdelinger i landet, en telefonrunde som det ofte kalles, for å få en oversikt over en viss hendelse. - Valgresultatene f.eks. blir etter hvert valgoppgjør ringt inn til NTB av valgstyrene når de er ferdig med opptellingen. - Markedsføringsfolk som vil "føle salget på pulsen" kan ha behov for kontinuerlig overvåking av utviklingen hos sine agenter, selgere e.l., og disse må da på bestemte datoer ringe inn sine siste salgsresultater.

Noen ganger kan det være en fordel å bruke telefon ved kontakt med oppgavegiverne etter at skjemaer er innsendt, enten for å rette opp feil eller for å få forklaringer til ting som kan synes uriktige. I Byrået brukes telefonen mye under revisjonen.

3. Intervjuundersøkelser

Intervjuundersøkelser er undersøkelser hvor opplysningene innhentes ved personlig intervju med et utvalgt antall personer som enten selv er enhetene eller sitter inne med opplysninger om enhetene. Slike undersøkelser får stadig større utbredelse.

Statistisk Sentralbyrå har bygget opp en egen intervjuorganisasjon fordi Byrået har intervjuundersøkelser med jamne mellomrom. Også visse private organisasjoner har egne intervjuorganisasjoner. Det kan skilles mellom intervjuere og tellere. Intervjuerne stiller spørsmål fra et skjema, mens tellerne kan ha som oppgave å levere ut/hente inn skjemaer og eventuelt også hjelpe til med utfyllingen. (I en del tilfelle kan imidlertid tellerne finne det mest hensiktsmessig å lese opp spørsmålene når de hjelper til med utfyllingen.)

Intervjuundersøkelser blir vanligvis mye dyrere pr. enhet enn postundersøkelser, fordi intervjuerne skal ha lønn og reiseutgifter dekket. Til gjengjeld vil en god tellerrepresentant kunne få mye bedre resultater kvalitetsmessig, enn en kan ved postundersøkelse. Vi vil dermed ved intervjuundersøkelser kunne spørre et mindre antall enheter enn ved postundersøkelser og likevel få et produkt som har like høy informasjonsverdi.

Intervjuundersøkelser bør be om oppgaver som kan gis på stående fot. Intervjueren kan ofte til en viss grad foreta en del kodinger allerede under intervjuet, og derved kan større bearbeidingsutgifter spares. Tellerne kan på stedet ut fra instruksjoner de har fått, korrigere eventuelle feil og misforståelser. Instruksene til intervjuerne kan ofte være meget detaljerte, og det er også tilfeller hvor intervjuerne er samlet til kurs for å kunne gjennomføre visse undersøkelser.

Vi har intervjuere når postkontakt med enhetene blir et ensidig og mangelfullt kommunikasjonsmiddel. Når det er mange og kompliserte spørsmål, vil telleenhetene kunne ha behov for utfyllende merknader etc. av en teller. En del telleenheter er dessuten ikke vant til skriftlig arbeid, og de må derfor hjelpes. Det er f.eks. ingen tvil om at intervjuundersøkelser er den eneste farbare vei ved de omfattende forbruksundersøkelsene. Forbruksundersøkelsen i 1973 omfattet 5 000 husholdninger som førte regnskap over sine utgifter og inntekter i en fjortendagers periode hver. For svært mange er det vanskelig å føre et regnskap, og derfor var det ofte helt nødvendig med personlig kontakt og veiledning. Veiledning ble gitt i tilknytning til et innledningsintervju, og intervjuerne tilbød hjelp underveis.

Det har videre vist seg at ikke bare kvaliteten av statistikken øker med gode intervjuere, men fracfallsprosenten reduseres også. Den dyktige, velinstruerte teller får færre avslag og bedre opplysninger enn den mindre dyktige og enn ved post. Sammenlikning med postundersøkelser vil likevel til en viss grad være avhengig av type spørsmål (visse "intime" spørsmål etc. besvares bedre ved post). Som en kuriositet kan nevnes at von Hofsten i boken "Praktisk statistikk" forteller at i Bolivia er det en statue over en folketeller som ble myrdet da han forsøkte å overtale en person til å gi opplysninger til en folketelling.

En stor fare ved intervjuundersøkelser er at alle intervjuerne ikke opptre likt til tross for en inngående instruksjon på forhånd. Dermed kan det ved resultatene oppstå et usikkerhetsmoment som en følge av intervjuernes subjektive instilling. Denne usikkerheten har vi ikke ved postundersøkelser fordi alle enhetene da får lik veiledning etc. Visse kontroller av intervjuernes arbeid tar sikte på å redusere denne usikkerheten.

Når vi skal telle enheter som ikke har adresse (trafikk, antall produserte enheter etc.), må vi ha tellere.

Intervjuerne bør som nevnt få en meget grundig instruksjon. Telleenhetene bør på forhånd informeres om eventuelle besøk av telleren, og telleren må ved besøk legitimere seg. Statistisk Sentralbyrå sender informasjon til politi og lensmenn i de distrikter undersøkelser skal foretas.

4. Registerundersøkelser

Ved registerundersøkelser ønsker vi å måle egenskaper ved enheter vi har samlet i et register. I virkeligheten er det en arkivstatistisk undersøkelse vi da snakker om. (Jfr. definisjonen på registre og arkiv under kapitlet om arkivstatistiske systemer.) Som regel vil disse registrene eller arkivene være opprettet for å tjene visse administrative formål. Denne form for undersøkelse omfatter derfor innsamling av data fra allerede innsamlet materiale som er lagret på en anvendelig måte. Statistikken

blir dermed et biprodukt av en administrativ prosess. Denne form for undersøkelser blir stadig mer utbredt. Dette skyldes i første rekke de nye tekniske hjelpemidlene, men også at slike undersøkelser på mange måter er fordelaktige. Vi er normalt sikret svar fra enhetene hvis registeret er i orden. Denne form for statistiske undersøkelser vil som regel også være billigere for Byrået hvis vi ikke inkluderer utgiftene med å opprette og ajourholde registeret. Hvis vårt register er opprettet og dessuten ajourholdes for andre formål (f.eks. administrative) enn en spesiell undersøkelse, bør de nevnte utgiftene verken helt eller delvis belastes den statistiske utnyttningen av dataene.

Registreringen kan skje sentralt eller desentralisert. Har vi en desentralisert registeroppbygging, vil man kunne få innsamlingsproblemer, idet vi da må hente inn opplysninger fra flere forskjellige registre.

Hvis man har de opplysningene man ønsker i et register, vil utarbeiding av statistikk i prinsippet kunne skje meget hurtig, og registerteknikken byr på meget store muligheter for statistikeren. Erfaringene viser imidlertid at det fremdeles er mange uløste tekniske og organisatoriske problemer; ikke minst koordinering av registre. I Byrået f.eks. har vi et bedrifts- og foretaksregister. Skattedirektøren opererer med et eget nummersystem for sine avgiftsenheter, selv om disse enhetene oftest faller sammen med Byråets foretaksenhet. Det er utarbeidd en nøkkel mellom disse to registrene, men dette var et svært omfattende arbeid. Nå ajourholdes Byråets bedrifts- og foretaksregister i første rekke gjennom meldinger til Skattedirektørens register.

Registre som ikke primært er opprettet for å tilfredsstille statistiske formål, vil neppe uten vanskeligheter direkte kunne nyttes til statistiske formål. Ofte vil statistisk utnytting av f.eks. administrative registre kreve planlegging, programmering og selvstendig maskinell bearbeiding (dvs. ikke som biprodukt av administrativ bearbeiding). Når vi skal gjennomføre en undersøkelse, bør vi alltid finne ut om det eksisterer et register hvor vi kan finne de enheter vi søker. Ofte gir ikke et register enhetene direkte, men vi ledes likevel til dem: Bedriftsregisteret inneholder f.eks. oppgaver over tallet på de ansatte i bedriftene. Det kan være til nytte ved en telling av lønnstakere som kan nåes via bedriftene slik det gjøres i lønnsstatistikkene. Via bedriftsregisteret kan vi gå til bedriftene og spørre om de ansatte, f.eks. deres lønn.

B. Blankettarbeidet - skjemaer og spørsmålsformulering

Skjemautforming rent teknisk vil avhenge av bearbeidings- og innsamlingsmetode. Spørsmålsformuleringen vil avhenge av problemstilling (hva vi skal spørre om), enhet (hvem vi skal spørre) og dessuten registreringsmetoden (hvordan vi skal spørre).

Mens vi formulerer og presiserer problemet, må vi skaffe oss et forenklet bilde av den virkelighet vi vil registrere, eller m.a.o. en modell. Vi skal registrere verdier på kjennemerker som beskriver enhetene i den forenklete verden vi vil undersøke. Man kan si at modellen, det forenklete bilde av virkeligheten, gir et sett av blankettopplysninger om enhetene, dvs. blanketten er faktisk vår modell. Derfor må den være hensiktsmessig utformet. Blanketten skal gi opplysninger fra statistikeren til telleenheten (rapporteringsenheten) og vice versa. Blanketten må derfor:

- 1) Identifisere utsender og enheter, dvs. sikre innhenting av data fra riktige enheter, om riktige enheter og opplyse hvem som ønsker disse dataene
- 2) Angi relevante kjennemerker, dvs. vi må få rede på det vi faktisk ønsker å få rede på, og ikke noe uvesentlig eller noe annet
- 3) Inneholde kontrollopplysninger, slik at vi er sikre på at 1) og 2) er riktig utført
- 4) Være lett å fylle ut
- 5) Være lett å bearbeide

Selv om vi lager en blankett som tilfredsstiller de fem kravene¹⁾ gitt foran, er det ikke dermed sikkert at blanketten er god. Det er mange måter å utforme en blankett eller et spørreskjema på, og

1) Ofte inndeles spørsmålene på et skjema i tre hovedtyper:

- a. Undersøkelsesspørsmål (jfr. krav 2)
- b. Bakgrunnsspørsmål (jfr. krav 1 og 2)
- c. Identifikasjonsspørsmål (jfr. krav 1 og 3)

Se f.eks. "Praktisk statistikk for samfunnsvidenskaberne" av K. Fisker, K. Nørgård, L. Thygesen, J. Wedeby og R. Østerlund, Akademisk Forlag (København 1977).

derfor bør skjemaet hvis mulig testes på forhånd, dvs. vi bør prøve skjemaet før vi lar oppgavegiverne komme i kontakt med det. En eventuell prøve kan omfatte flere deler av en undersøkelse enn akkurat skjemaet (f.eks. kan bearbeidingsrutinene testes).

Vi skal så gå litt nærmere inn på de fem krav vi stilte til blanketten.

Et spørreskjema må alltid inneholde opplysninger om avsender og hans adresse (inkl. telefon). Spørreskjema bør benevnes stikkordmessig, dvs. gis et kort navn, f.eks. Industristatistikk 1976 (se figur 6). Vi må så tidlig som mulig i skjemaet opplyse om hvilken periode eller tidspunkt undersøkelsen omfatter, og når vi vil ha skjemaet utfylt, dvs. tidsfrist.

Figur 6

Statistisk Sentralbyrå
5. kontor
Dronningensgt. 16, Oslo
Postboks 8131 Dep., Oslo 1
Tlf. (02) 41 38 20

Industristatistikk 1976

Undergitt taushetsplikt

Frist for innsending
10. mars 1977

33112, 33114, 33119, 3312, 3319, 332

Les nøye gjennom rettleidingen for utfylling av skjemaet

Kontoradresse

I de aller fleste tilfeller skal de innsendte oppgavene være undergitt taushetsplikt, og derfor bør dette påføres klart og tydelig, slik at ikke enkelte oppgavegivere lar være å svare fordi de er redd andre enn statistikerne skal få rede på det de oppgir. Utsenderen bør også passe på å oppgi om det er en lovhjemmel for den aktuelle undersøkelsen eller om den er basert på frivillighet.

Særlig hvis undersøkelsen ikke er obligatorisk, dvs. hjemlet ved lov, må oppgavegiverne motiveres til å besvare skjemaet. Det bør ved skjemaet til alle undersøkelser også ligge et følgebrev som sier noe om formålet med undersøkelsen, hva den skal søke å gi svar på, slik at oppgavegiverne forstår at det er viktig at de besvarer spørsmålene i skjemaet. Foruten motivering kan et slikt følgebrev inneholde en rettledning for oppgavegiverne. Rent fysisk bør av og til motivasjonen og rettledningen være to atskilte brev eller blanketter, slik at vi i alt kan få tre blanketter, hvorav bare én - nemlig skjemaet - inneholder spørsmål og skal utfylles. Rettledningen kan også av og til være trykket sammen med spørsmålene, dvs. bare en blankett. Selv om statistikeren konstruerer et skjema med atskilt veiledning, må han huske på at en veiledning sjelden leses; spørsmålene må derfor være klare og tydelige.

Vi har av to grunner behov for å identifisere enhetene vi skal spørre: Statistikeren må være sikker på at han får oppgave fra den enhet han ønsker, og oppgavegiveren må forstå at spørsmålene gjelder nettopp ham og at ikke blanketten(e) bare ved en feiltakelse har havnet hos ham. Hvis oppgavegiveren ikke er enheten vi ønsker opplysninger om, må dette også presiseres tydelig (f.eks. oppgavegiver er familiefaren, mens enheten kan være familien). Selv om vi ved utsendelse av skjemaene har påført navn og adresse på oppgavegiveren, vil vi oftest be oppgavegiveren gi opplysning om navn og adresse. Vi kan da få kontrollert og eventuelt rettet opp de identifiseringsopplysningene vi sitter inne med. (Som vi husker samler vi identifiseringsopplysningene i registre, og dermed kan vi altså få ajourført visse registre.)

Spørsmålene bør komme i en logisk rekkefølge (hvis vi f.eks. ønsker sysselsettingen i alle årets måneder og gjennomsnitt for året, bør vi først be om sysselsettingen i de enkelte månedene og så gjennomsnittet, ikke omvendt, selv om det kanskje er gjennomsnittet vi først og fremst er interessert i).

Spørsmålene kan være utformet slik at svaralternativene enten er åpne eller lukkede (bundne). Et lukket eller bundet svaralternativ vil si at det er angitt et begrenset antall muligheter for svar, f.eks. ja eller nei; gift, ikke gift, osv. (se figur 7, hvor spørsmålene 1 og 3 er spørsmål med bundne svaralternativer). Ofte har man da også med et svaralternativ som heter "annet" (dvs. annet enn angitt i alternativene), men da ber en også om en spesifisering fordi "annet" ofte kan grupperes under et av de andre angitte alternativene (se spørsmål 9 i figur 7). Et åpent svaralternativ vil si at oppgavegiveren kan svare nesten hva som helst (se f.eks. spørsmål 2 i figur 7). Åpne spørsmål vil som regel ta mindre plass på skjemaet enn lukkede og tillate svareren å gi uttrykk for flere nyanser i sitt svar enn et lukket spørsmål uten "åpen ende" (dvs. "annet" som siste alternativ).

Avgrensingen av det forhold man er interessert i (den presise definisjonen av kjennemerket) kan være svært vanskelig, men er desto viktigere: Hva er f.eks. en bolig, en leilighet, en hybel? Hva er en butikk (inngår kontor og lagerrom)? Ofte kan det ved spørsmål om holdninger (f.eks. til norsk u-hjelp) være nødvendig å gi visse premisser eller forutsetninger. (Man kan være for økt u-hjelp så lenge "andre" må betale regningen, og det kan man kanskje regne med at intervjuobjektet tror hvis dette ikke blir presisert i spørsmålsstillingen?)

Vanskelighetsgraden i spørsmålene bør være avpasset oppgavegiverne, og en bør spørre bare om en ting ad gangen. Spørsmålene bør være "lette" å forstå og svare på. I desember 1976 stilte Gallup spørsmål om demping av skatteprogresjonen til et representativt utvalg på 800 personer over 15 år (Aftenposten 26/2-77). Etter en forklaring på hva det ville bety for statskassen om ingen betalte mer enn 50 prosent av sin inntekt i skatt, ble de intervjuede spurt om de var for en slik demping. Det er tvilsomt om alle forstår hva spørsmålet innebærer, og det er også tvilsomt om det er den statsfinansielle siden av progresjonen som er det sentrale. (Fra politisk hold er det inntektsutjamningssiden som har vært trukket mest fram.) Ved personlig besøk, dvs. intervju, kan det riktignok stilles vanskeligere spørsmål fordi intervjueren da kan oppklare vanskeligheter på stedet. Vi må også huske at det er visse grenser for hva folk vil oppgi om seg selv, enten det dreier seg om svar pr. post eller ved intervju. Statistikk over alkoholkonsumet, innhentet ved hjelp av spørsmål til konsumentene, viser vesentlig lavere tall enn Vinmonopolets tall. Hva som er "følsomme" spørsmål, forandrer seg med tiden. For noen år siden var spørsmål om inntekt ganske ømtålig, i dag er spørsmål om utdanning mer ømtålig (det er ikke "fint" bare å ha grunnskoleutdanning, selv om 2/3 av befolkningen over 16 år ikke har høyere utdanning). Også svarene til de politiske gallupmålingene tyder på at folk enten glemmer svært lett eller ikke vil stå ved det de stemte på sist (kanskje det ikke er "fint" nok).

Figur 7.

STATISTISK SENTRALBYRÅ
Underavdelingen for
intervjuundersøkelser
Postboks 8131 Dep., Oslo 1
Tlf. *(02) 41 38 20

Intervjuerens navn:

Intervjuerens nr.:

ARBEIDSKRAFTUNDERSØKELSEN
4. kvartal 1976

UNDERGITT TAUSHETSPLIKT

Skjematype

Utv. omr. nr.

Ar-Kvartal-Gang

Husholdningsnr.

			1	1
				2- 4
7	6	4		5- 8
				9-12

Navn og adresse		Fødselsdag-mnd.-år		Personnr.	
				13-23	
1. Utførte De inntekts-givende arbeid av minst en times varighet i forrige uke?		24	1 ☐ Ja	→ 2	
		2 ☐ Nei		→ 3	
2. Hvor mange timer arbeidet De i forrige uke? Ta med overtidarbeid og ekstraarbeid, også ekstraarbeid hjemme i forbindelse med arbeidet.		25-26	Under 21 timer	→ 7	
		☐	21 timer og over	→ 21 a	
3. Har De inntekts-givende arbeid som De var borte fra i forrige uke?		27	1 ☐ Ja	→ 8	
		2 ☐ Nei		→ 4	
4. Er De selv eller noen i Deres husholdning selvstendig yrkesutøver (f.eks. gårdbruker eller eier av egen tikk eller verk)		28	1 ☐ Ja	→ 5	
		2 ☐ Nei		→ 9	
5. Utførte De					
9. Hva gjorde De hovedsakelig i forrige uke?		34	1 ☐ Var syk	→ 14	
		2 ☐ Utførte husarbeid hjemme		→ 10	
		3 ☐ Gikk på skole/studerte		→ 10	
		4 ☐ Var inne til l. gangs militær- eller sivilarbeidstjeneste		→ 14	
		5 ☐ Var arbeidsufør		→ 14	
		6 ☐ Var pensjonist/sluttet i arbeid		→ 10	
		7 ☐ Var uten arbeid		→ 10	
		8 ☐ Opptatt med annet (spesifiser):		→ 10	
10. Forsøkte De å få inntektsgivende arbeid for forrige uke? Vi tenker her på forsøk som ble gjort for å få arbeid som De kunne ha påtatt Dem i forrige uke.		35	1 ☐ Ja	→ 11a	
		2 ☐ Nei		→ 14	

Spørsmålene må ikke innby til tendensiøse svar. Husk at de fleste mennesker ønsker å være "bedre" i en eller annen forstand enn de faktisk er. Derfor vil spørsmål som går ut på hva enheten "ønsker", besvares helt annerledes enn hva han "har gjort" eller faktisk "vil gjøre".¹⁾ Av samme grunn bør også bestemte verdiladede begreper som profitt, sosialisering, borgerlig, byråkratisk, kapitalist, etc. unngås.

Spørsmålene må ikke omfatte mer enn en egenskap om gangen, men av og til kan dette være vanskelig å få til.

1) Drammens Tidende og Buskeruds Blad kan f.eks. fortelle at ved spørsmål har en meget stor andel av leserne oppgitt at de leser avisens leder. Kontrollspørsmål som har gått på innholdet i lederen har vist at langt færre har lest lederen.

Selve rekkefølgen av spørsmålene vil kunne ha betydning. Et eksempel fra en undersøkelse som Gallup utførte i 1976 viser dette¹⁾. Spørsmålet som ble stilt var: "Mener De at levestandarden i Norge stort sett er altfor høy, noe for høy, passende, noe for lav eller altfor lav?". Dette spørsmålet ble stilt som nr. 30 av i alt 70 spørsmål. Det etterfulgte spørsmål om spisevaner, matvarer og matvareproduksjon, kostholdsplanlegging, helse, selvberging og u-landsproblemer. Aftenposten sier i sin kommentar at spørsmålet er farvet av disse innledende spørsmål. For å underbygge denne påstanden lot Aftenposten Gallup stille det samme spørsmål, men da frittstående uten innledningsspørsmål. Svarfordelingen ved de to undersøkelsene er gitt nedenfor i tabell 1. I den siste undersøkelsen ble det også stilt spørsmål om egen levestandard ("Og hva med Deres egen levestandard - mener De den stort sett er altfor høy, noe for høy, passende, noe for lav eller altfor lav?"). Svarene kan tyde på at mange folk mener at det er andres levestandard som er for høy og ikke deres egen.

Tabell 1. Gallup om levestandard

	Levestandarden i Norge		Egen levestandard
	Svar farvet av innledende spørsmål	Frittstående spørsmål	(Spm. etter frittstående spm. om levestandard i Norge)
		Pst.	
Sum	100	100	100
Altfor høy	28	15	3
Noe for høy	48	43	23
Passende	22	38	66
Noe for lav	1	2	7
Altfor lav	0	0	1
Vet ikke	1	2	1

K i l d e: Aftenposten 14/2-76. Undersøkelsen omfattet ca. 1 600 personer over 15 år.

Når vi utarbeider et nytt skjema, bør vi ta sikte på å beholde det uforandret i noen år. Det blir både billigere og tidssparende for statistikkprodusenten, samtidig som utfyllerne vet hva de har i vente og kan innrette seg deretter, f.eks. legge opp sine regnskaper etter de spørsmål de vet må besvares. Dette synspunktet gjelder selvfølgelig bare løpende undersøkelser, men likevel kan et tilsvarende råd gis når det gjelder engangsundersøkelser: Det er en fordel å bruke standardiserte formuleringer og betegnelser i spørsmål som ofte går igjen, f.eks. i bakgrunnsspørsmål (yrke, sivilstand, helse o.l.).

Antall spørsmål eller lengden på skjemaet bør også vurderes nøye. Man bør bare ta med de spørsmålene som virkelig vil ha betydning for undersøkelsen og så sløyfe spørsmål som "det kunne være interessant å ha med", fordi kvaliteten på svarene ofte synker med størrelsen på skjemaet. Og det er lite ønskelig å få rede på noe som er på kanten av vår interesse hvis det samtidig går ut over kvaliteten i det vi virkelig vil undersøke. Spesielt når man har et arkivstatistisk system bør man undersøke nøye hvorvidt en del opplysninger om de enhetene man vil undersøke, like godt kan hentes fra andre undersøkelser. Oppgavegiverne blir ofte irritert over å oppgi samme ting på to forskjellige skjema, men til samme institusjon. De har også en tendens til å oppfatte staten som en og samme institusjon, dessuten synes det å være liten forståelse for at et skjema skal dekke over mange individuelle variasjoner slik at det egentlig ikke passer "skreddersydd" for noen, men bare tilnærmet for alle.

1) Aftenposten 14/2-76.

Når det gjelder selve den tekniske siden av skjemaet, må vi ta hensyn til både utfyller og arbeider. Skjema til bedrifter o.l. må kunne utfylles med skrivemaskin. Skal et skjema kunne utfylles med maskin, må vi ta hensyn til skrivemaskinstandarder. Det må være plass nok til alle alternative svar, ellers vil oppgavegiveren kunne tro at hans svar er galt.

Papirkvaliteten betyr mye og må tilpasses utfyllingsmåten. Blankt papir kan være vanskelig å lese (refleks).

Er det flere typer skjema som sendes ut av samme institusjon eller kontor og som bærer samme navn, kan et skille i farge på skjemaene være gunstig: Alle små industribedrifter får "sorte" skjemaer (dvs. sort skrift), mens alle de store får "grønne" skjemaer. Dette vil lette eventuelle kontakter med enhetene og forenkle sorteringer av de innsendte skjemaene.

Spørsmålene bør fortrinnsvis nummereres, og nummereringen må være hensiktsmessig både for utfylling, bearbeiding og punching.

Ofte er det spesialister som arbeider med den tekniske siden av skjemaautformingen, fordi en har kommet til at et godt teknisk tilpasset skjema betyr utrolig mye for kvaliteten på de svarene en ønsker å få.

5. BEARBEIDING AV OBSERVASJONSMATERIALET

A. Revisjon, kontroll og oppretting

De innhentede data kan enten revideres manuelt eller maskinelt, altså revisjon ved hjelp av programmerbare EDB-maskiner, eller en kombinasjon av de to metodene. Også data som "innhentes" fra registre eller administrative rutiner må revideres¹⁾. Som regel vil dataene bli innhentet på skjemaer, men framgangsmåten angitt nedenfor kan også brukes på data innhentet via magnetbånd o.l. Under planleggingen av en statistisk undersøkelse må vi avgjøre om de innsendte oppgavene skal revideres manuelt eller maskinelt. Noe skarpt skille mellom de to revisjonsmåtene er det ikke i dag, ofte er det fordelaktig med en integrert maskinell og manuell revisjon. Når skjemaet er fastlagt og vi vet hvilke formål undersøkelsen skal tilfredsstille, kan vi utarbeide revisjonsinstrukser, dvs. bestemme oss for hvordan dataene skal revideres.

Ved revisjon av statistiske data innhentet på skjema, må man se sammenhengen mellom spørsmålene, huske på at skjemaet gir et forenklet bilde av virkeligheten og at det derfor er en sammenheng mellom spørsmålene og at det derfor også er det mellom svarene, dataene.

Revisjon av statistiske data vil si kontroll og oppretting av de data oppgavegiverne har gitt. Det bør alltid utarbeides en instruks for hvordan dataene skal revideres. I revisjonsinstrukser bør det være gitt retningslinjer for: 1) Hvilke kjennemerker som skal kontrolleres, 2) Hvilke verdier kjennmerkene kan tillates å ha (altså hvilke data som skal kunne forekomme) og 3) Hvordan en eventuell oppretting av gale eller ufullstendige data skal foretas. - Under utarbeidelsen av revisjonsinstrukser må man derfor "tenke gjennom" hele undersøkelsen og finne fram til mulige feil som kan oppstå. Revisjonsinstrukser utarbeides for å:

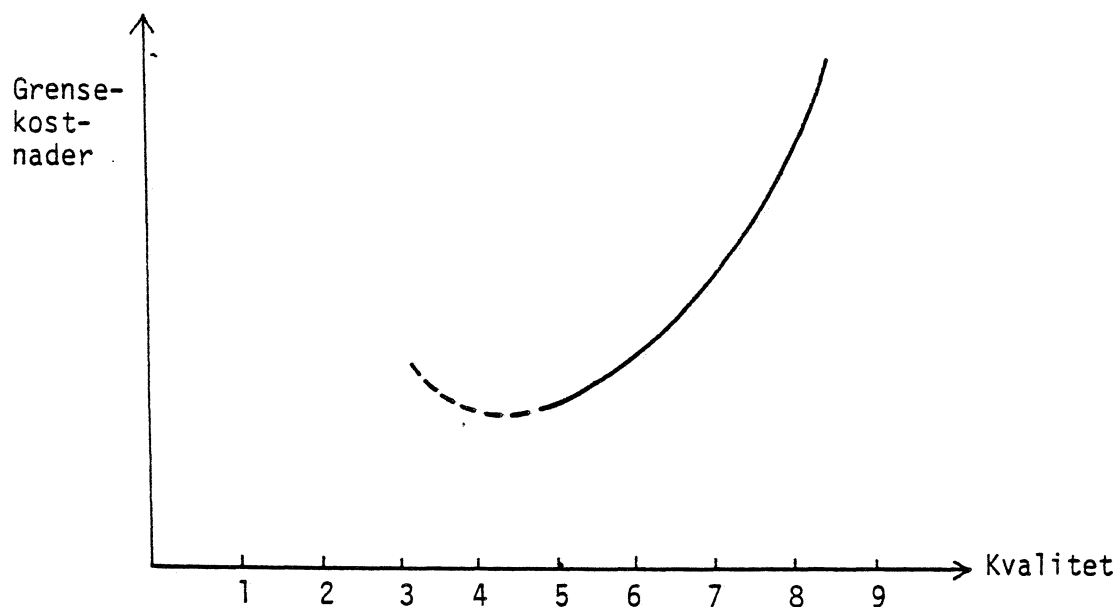
- 1) Sikre en enhetlig revisjon av dataene (slik at den enkelte revisors egne ikke-faglige vurderinger ikke "farger" revisjonen, men selvfølgelig må revisors faglige vurderinger komme til nytte i revisjonen)
- 2) Sikre at hvis en revisor slutter i jobben, skal dette ikke ha betydning for det endelige resultat
- 3) Sikre at revisorene ikke blir glemske etter hvert
- 4) Lære opp nye revisorer

1) Imidlertid vil registerdata sjelden revideres på samme måte som data innsamlet på vanlige statistiske skjema. (Se f.eks. Arbeidsnotater IO 79/1 av Tor Nygaard: Det sentrale bedrifts- og foretaksregisteret. Innhold og muligheter. Spesielt siste avsnitt i punkt 5.1).

Revisjonen kan være oppdelt i flere trinn, slik at en utnytter den enkeltes kvalifikasjoner best mulig. Noen er trent til å summere raskt og sikkert, andre har utdanning og erfaring i å lete etter feil eller mangler. Noen feilsøkinger passer best for EDB-maskiner, mens andre må foretas av mennesker, ellers vil revisjonsprogrammene måtte gjøres meget omfattende (og dermed bli meget kostbare både i konstruksjon og drift).

Også revisjon av dataene må sees i sammenheng med undersøkelsens andre ledd. Det er ikke lønnsomt å finrevidere data hvis det endelige statistikkprodukt ville fått en høyere informasjonsverdi ved å bruke "finrevisjonspengene" på bedre planlegging isteden. Igjen må planleggeren søke å avveie utbytte og kostnader ved de enkelte ledd i undersøkelsen. Også når det gjelder revisjon, må vi regne med stigende grensekostnader, slik som illustrert i figur 8. Det koster atskillig mer å øke kvaliteten fra 7 til 8 enn fra 5 til 6 (dette kan f.eks. tolkes slik: Det koster mer å redusere antall målefeil fra 20 til 10 prosent enn fra 40 til 30 prosent). Som regel er det meget kostbart å revidere et materiale så grundig at det blir helt feilfritt.

Figur 8



Revisjon vil si å kontrollere så langt som mulig (økonomisk og/eller teknisk) at vi har fått inn korrekte oppgaver. Revisjonen kan gjøres både sentralt og lokalt. Vi har to hovedgrupper av revisjon:

1. Oppgavekontroll og blankettkontroll
2. Resultatkontroll

Oppgavekontrollen går ut på å se om enhetene har besvart spørsmålene som er stilt dem, eller om de har svart på gale spørsmål. Vi må forsøke å rette opp slike gale opplysninger.

Blankettkontrollen går ut på å se om vi har fått inn skjemaer fra de riktige enhetene, og om enhetene er rett identifisert på skjemaene. Dette er som regel en kontroll av om slippdataene (dataene på navne- og adresseslippene påklisset skjemaet ved utsendelsen), stemmer overens med identifikasjonsopplysningene oppgavegiver er bedt om å gi, dette er en form for register- eller arkivkontroll.

Resultatkontroll vil si at vi kontrollerer (vurderer) resultatene i de tabellene vi kommer fram til.

Kontrollarbeidet kan vi dele i to trinn:

- a. Identifisere gale og tvilsomme opplysninger
- b. Etterkontroll og retting av gale og tvilsomme oppgaver

Generelt kan vi regne med tre typer av kontroller:

- a. Kvalitetskontroll
- b. Prosesskontroll og
- c. Innbyrdes kontroll av oppgavene

Kvalitetskontroll vil si at vi undersøker unøyaktige, dårlig utfylte, mangelfulle og liknende svar. Prosesskontroll vil si at vi kontrollerer definisjonsmessige logiske og sannsynlige sammenhenger i de svarene vi har fått, f.eks. om et visst antall tonn råolje er blitt raffinert til et rimelig antall liter bensin, fyringsolje etc. Alder og antall barn kontrolleres mot hverandre, f.eks. har en attenåring ikke fem barn, osv. Også sumkontroller faller inn under dette punkt. Når de enkelte svarene på et skjema blir stilt opp mot hverandre, kalles det gjerne for intern konsistenskontroll.

Innbyrdes kontroll av oppgavene går ut på å kontrollere svarene som enhetene har gitt, slik at en enhets svar blir kontrollert mot andre enheters svar og mot eventuelle tidligere undersøkelser. Dette kalles også kontroll av massekonsistens.

Hva som bør gjøres maskinelt og hva som bør gjøres manuelt av disse leddene, vil avhenge av selve undersøkelsen: Revisjon som krever lite skjønn, bør gjøres maskinelt. Revisjon som går ut på å kontrollere en enhets svar mot andre enheters svar (dvs. mot gjennomsnittssvar), bør gjøres maskinelt. Revisjon som krever spesialinnsikt, må foretas manuelt. Resultatkontroller og sumkontroller kan best foretas maskinelt. Som vi skal se seinere, kan også ulike former for etterkontroller og kontroller av revisorene med fordel utføres maskinelt.

På side 54 er listet fire punkter for hvorfor det utarbeides revisjonsinstrukser (tilsvarende for koding)¹⁾. Alle punktene som er listet der, bunner ut i formålet: Sikre en så god og standardisert revisjon som mulig (jfr. punktene 1-4). En revisjonsinstruks bør imidlertid også tilfredsstille flere formål. Den skal: 1) Dokumentere hva som er gjort med de innsamlede rådata, 2) Angi hvilken vekt revisorene skal legge på de enkelte deler (enheter og/eller spørsmål) av materialet, 3) Angi ansvarsforhold, tidsramme og hjelpemidler som kan (skal) brukes.

Innholdet i en revisjonsinstruks

En revisjonsinstruks bør inneholde - gjerne starte med - en slik oversikt av generell karakter, men også gi eksempler på hva de ulike kontrolltyper vil si for det bestemte datamaterialet instruksen er utarbeidd for.

1) Koding er nærmere omtalt på side 60.

Ofte kan det være hensiktsmessig å dele revisjonen opp i flere trinn, f.eks. i tre:

1. Identifisere gale og tvilsomme opplysninger (jfr. gruppe 1, trinn a på side 56). Denne delen av arbeidet kan ofte med fordel også deles i to. Den første delen vil da kunne være en innsjekkingskontroll, dvs. kvittering o.l. for innkomne, godtagbare skjemaer
2. Etterkontroll og retting av gale og tvilsomme oppgaver (jfr. gruppe 1, trinn b)
3. Resultatkontroll (jfr. gruppe 2)

Hensikten med å dele revisjonen opp i flere trinn, vil først og fremst være å få frigjort den kvalifiserte arbeidskraften fra mer trivielle arbeidsoppgaver til vanskeligere revisjonsarbeid.

Hvis vi deler revisjonen opp i flere trinn, kan vi også utforme revisjonsinstrukser for hvert trinn. Disse instruksene kan da bedre tilpasses kunnskapene til revisorene på de enkelte revisjonstrinn, men selvfølgelig bør helheten i revisjonen framgå for alle revisorene.

En revisjonsinstruks bør, enten den er hel eller delt i flere trinn, inneholde følgende deler:

- (1) Feilsøkingsdel. Instruksjonen bør angi hvilke typer av feil som kan forekomme, helst med angivelse av hyppighet (dvs. hvor ofte feilen forekommer). Instruksjonen bør iallfall gi konkrete eksempler på feil typer, men helst en uttømmende liste over mulige feil. Det kan ofte være hensiktsmessig på dette trinn av revisjonen å operere med et revisjonsark som angir en rekke kontroller revisorene skal utføre og visse spørsmål som skal besvares. Feilsøkingen kan også være delt opp i flere deler, en lavere del (enkle feil) og en høyere del.

Hvordan skal vi så finne ut om svaret på et spørsmål er galt, eller iallfall tvilsomt? Da må vi bygge på det vi vet eller får vite under revisjonen om det enkelte svars sammenheng med andre opplysninger vi har om enheten eller hele massen. Vi utfører det vi kaller konsistenskontroller, som vi har tre stykker av:

- a) Intern konsistens: Vi sammenlikner alle svarene på blanketten med hverandre, f.eks. om svare på det første spørsmålet er forenlig med seinere spørsmål
- b) Register-, eller arkivkonsistens: Vi sammenlikner blankettens opplysninger (svarene) med andre opplysninger vi har om enheten, f.eks. fra registre eller fra arkiv (dvs. fra blanketter fra tidligere eller andre tellinger)
- c) Massekonsistens: Vi sammenlikner enhetene med hverandre, dvs. om det er sammenheng mellom svarene som er gitt for den enkelte enhet og svarene som er gitt for massen totalt

En revisjonsinstruks må inneholde anvisninger på hvordan revisor skal komme på sporet etter inkonsistenser (dvs. brudd på a, b eller c ovenfor). Anvisningene bør ofte gis i form av at instruksjonen påpeker og opplyser om logiske og definisjonsmessige sammenhenger. Er ikke disse sammenhengene tilstede, er det trolig en inkonsistens.

Eksempler på logiske sammenhenger er: Ingen 10-åring er gift; ingen personlig eier av et foretak hvis navnet angir et akseselskap; summen er like stor som underpostene dersom alle underpostene er fullstendig spesifisert; et foretak bør betale gjeldsrenter hvis det har gjeld, osv.

Av definisjonsmessige sammenhenger har vi to typer: fysiske og strukturelle. Eksempler på fysiske sammenhenger: Vekten av et produkt kan ikke avvike vesentlig fra den samlede vekt av råstoffene; visse råstoffer er nødvendig for å produsere bestemte produkter, osv. Strukturelle sammenhenger er sammenhenger vi ofte kommer fram til på grunnlag av empiriske (tallmessige) målinger av enhetene og egenskapene. Vi vil dermed kunne få skilt ut "unormale" tilfelle, f.eks. at familier ikke bruker mindre enn 20 prosent eller mer enn 50 prosent av sin inntekt til mat.

- 2) Opprettingsdelen. På grunnlag av feilsøkingen kan vi plukke ut en rekke svar som åpenbart er gale og mange som sannsynligvis er gale. En rekke svar kan rettes umiddelbart eller etter kontakt med oppgavegiver. Når det gjelder oppretting, hvis oppgavegiver ikke kan kontaktes, har vi følgende muligheter:

- a) Akseptere de gale eller tvilsomme svarene
- b) Se bort fra de manglende og feilaktige svarene
- c) Se bort fra enhetene som på ett eller flere punkter har gitt tvilsomme eller feilaktige oppgaver
- d) Rette etter skjønn

Instruksen bør angi hvordan de forskjellige typer av feil skal rettes. Det bør angis når oppgavegiver skal kontaktes, og det bør angis hvordan man skal rette opp feil i de tilfelle da oppgavegiver ikke behøver kontaktes. På dette trinn av revisjonen vil en da bruke det eventuelt utfylte revisjonsarket, en vil bruke det under opprettingen.

I instruksen bør ansvarsforholdet være klart angitt, hvem som har ansvar for de enkelte deler av revisjonen, hvem som eventuelt skal kontakte oppgavegiverne og hvordan kontaktbrevene (revisjonsbrevene) skal (bør) utformes.

Opprettingen er en vanskelig del av revisjonen, og derfor vil revisorene ofte på dette trinn ha behov for så mange hjelpemidler som mulig: Skjema som tidligere er innsendt, litteratur, annen statistikk, avis- og tidsskriftutklipp, årsberetninger osv. I instruksen bør det være gitt henvisninger til slike hjelpemidler.

I revisjonsinstruks må alltid angis hvordan opprettinger skal foretas rent teknisk: Med rød eller grønn blyant, ikke utstrykninger av oppgavegivers svar, osv.

Når det gjelder feilsøking og oppretting, er det viktig å eliminere feil som a) Er store i forhold til totalen, b) Gjelder store enheter og c) Gir asymmetriske utslag (gale statistiske fordelinger).

(3) Kodeliste - stikkordregister. En instruks bør være lett å finne fram i, også for en lite øvet revisor. Foruten kodeliste(r) bør derfor instruksene helst inneholde stikkordregister, spesielt gjelder dette hvis instruksene er delt opp i flere deler eller trinn, slik at samme ting (feil, emne etc.) kan være omtalt flere steder.

En instruks bør også inneholde en innholdsfortegnelse, og den bør ha daterte sider som kan innsettes i en ringperm e.l. hvis den skal brukes for flere årganger. Har man daterte sider, vil det være lett å skifte ut, foreta forandringer.

(4) Bearbeidings- og resultatkontroll. En revisjonsinstruks må ikke bare inneholde kontroller for de enkelte enheter eller svarene som de enkelte enheter gir, men også gi anvisninger på hvordan hele bearbeidningen og resultatene skal kontrolleres.

- a) De sammenhenger som gjelder for den enkelte enhet bør i de fleste tilfelle også gjelde for hele massen, eller delmasser. Vi bør se på resultatene for hele massen i relasjon til tidligere undersøkelser
- b) Bearbeidingsprosessen må kontrolleres fordi det oppstår feil overalt hvor mennesker behandler store mengder tall. Kontrollen bør finne fram til feil, slik at bearbeidningen kan bli bedre seinere, eller at spesialinstruks e.l. kan gis til "dårlige" revisorer
- c) Utfyllingen bør hvis mulig kontrolleres i noen grad, dvs. vi bør iallfall av og til gå tilbake til enhetene for å finne ut om de virkelig gir oss opplysningene vi ber om, eller om de har misforstått. Dette vil være kostbart, og det er derfor bare aktuelt ved særlig viktige undersøkelser

(5) Tidsplan. En revisjonsinstruks bør inneholde en tidsplan for revisjonen, slik at de enkelte revisorene forstår hvor lang tid revisjonen skal/kan ta.

I tidsplanen bør også angis hva det skal legges vekt på under revisjonen, hva som vil føre til betydelige feil i det endelige produkt og hva som har ubetydelig interesse. Ved revisjon av industristatistikksskjemaer f.eks. er det klart at A/S Norsk Jernverk bør tillegges mer vekt og dermed ta lengre revisjonstid enn en liten skofabrikk (med f.eks. 20 ansatte), selv om de har fått tilsendt like store skjema og skjemaene er like "dårlig" utfylt. Vi kan godta flere feil hos enheter som betyr lite i motsetning til enheter som betyr mye i den endelige statistikken. "Mye" og "lite" må her sees i forhold til spesifikasjonsnivået ved offentliggjøring, f.eks. om det skal gis kommunedata.

(6) Eksempler. Ofte vil det være nødvendig å angi detaljerte eksempler i en eventuell instruks på hvordan revisjonen skal foregå. Hvorvidt eksempler skal gis i selve instruksene eller under en opplæring i bruk av instruksene, vil ofte være et spørsmål om hensiktsmessighet.

1. Manuell bearbeiding kontra EDB-bearbeiding

Følgende seks faktorer har betydning når vi for én bestemt undersøkelse skal velge mellom maskinell (EDB-maskiner) og manuell bearbeiding (inkl. revisjon):

- 1) Kostnadene for bearbeiding av materialet for hver av de to bearbeidingsmetodene
- 2) Hastigheten som materialet kan bearbeides med etter de to metodene
- 3) Vanskelighetsgraden i bearbeidingen, om det er enkle eller kompliserte tabeller, f.eks. såkalte krysstabeller, dvs. tabeller som først kan lages ved ulike sorteringer av datagrunnlaget. EDB-maskiner er spesielt egnet til å sortere et datamateriale hurtig og billig
- 4) Kapasiteten hos personalet (bearbeiderne) og EDB-maskinene
- 5) Mulighetene for maskinell revisjon samtidig med at materialet bearbeides (taloppstillinger mv.)
- 6) Mulighetene for videre maskinell bearbeiding av dataene på et seinere tidspunkt i forbindelse med annet statistisk materiale; jfr. arkivstatistiske systemer

Ønsker vi å opprette et arkivstatistisk system, må vi overføre dataene på et eller annet tidspunkt til en maskinlesbar form (f.eks. magnetbånd). Dette behøver ikke bety at den aktuelle undersøkelsen bør bearbeides maskinelt fordi EDB-bearbeiding kan ta for lang tid i forhold til manuell bearbeiding (det tar tid å utarbeide EDB-programmer m.m.), men som oftest vil en da med en gang nytte EDB-maskin.

Et statistisk materiale bearbeides i praksis både manuelt og maskinelt, ofte utarbeides de endelige tabeller ved hjelp av bordregnemaskiner på grunnlag av EDB-lister (maskinlister). I den seinere tid har imidlertid stadig flere tabeller beregnet for publisering (trykking) blitt skrevet direkte ut av EDB-maskinen på offsetoriginaler (såkalt hvitt papir). Dermed unngår en også mulige overføringsfeil, dvs. feil ved overføring av tall fra EDB-lister til tabeller eller skrivefeil ved vanlig manusskriving. Visse manuelle kontroller vil det også være nødvendig å foreta etter en eventuell maskinell bearbeiding, men det skal vi komme nærmere inn på seinere.

Kostnadene ved bearbeidingen spiller en stor rolle. Ved maskinell bearbeiding har vi tre hovedgrupper av kostnader: 1) Maskinkostnader, 2) Programmerings- og systemkostnader og 3) Gjennomførings-, eller driftskostnader. Ved manuell bearbeiding har vi ikke kostnadsgruppe 2); dessuten blir selve maskinkostnadene helt ubetydelige (fordi det bare er de variable kostnader som har betydning i denne sammenheng, og variable kostnader (slitasje) er svært små for manuelt opererte maskiner). Til gjengjeld blir driftskostnadene (lønn til regnemaskinoperatører mv.) som regel betydelig større ved manuell bearbeiding. Det utstyr som må være tilstede for en maskinell bearbeiding deles ofte opp i "Hardware" og "Software". Hardware står for EDB-maskiner, mens Software står for programmer, systemer o.l. Ofte kan samme Software brukes for flere forskjellige bearbeidinger, og dermed vil kostnadene kunne fordeles på flere kostnadsbærere. Dette vil igjen si at når man skal velge mellom maskinell og manuell bearbeiding, må en ikke bare ha oversikt over det maskinelle utstyret som finnes, men også over hvilke systemer og programmer som foreligger ferdig til bruk, dvs. om det må spesialprogrammering til eller om det allerede finnes programmer utarbeidd som er såpass generelle i sin form at de kan brukes på den aktuelle undersøkelsen. Brukes generelle programmer eller standardprogrammer vil selvfølgelig Software-kostnadene kunne bli meget små (programmeringskostnadene kan da fordeles på flere prosjekter), mens derimot Hardware-kostnadene som oftest blir større enn ved spesialprogrammering (fordi program som skal kunne brukes for flere enn et prosjekt som regel må være større enn nødvendig for det ene prosjektet). Hittil har prisen blitt stadig lavere på Hardware og stadig høyere på Software. Denne utviklingen kan vi også vente i tiden framover. Derfor vil et riktig valg av bearbeidingsmetode i dag ikke nødvendigvis være riktig i morgen.¹⁾

1) IBM oppgir f.eks. til Aftenposten 15/2-79 at det på et utvalg av IBM's sentralenheter fra 1973 til 1979 har vært en forbedring i forholdet pris/ytelse på 7 ganger. Omtrent samme utvikling har det vært på de andre maskinene som inngår i et EDB-system. Produksjonskostnadene er f.eks. redusert til en fjerdedel av hva de var i 1971. IBM opplyser samtidig at kostnadene ved å utvikle systemer og programmer stadig øker. Om forholdet for 10 år siden var 70 prosent av EDB-kostnadene til maskinelt utstyr og 30 prosent til utvikling, viser erfaringene i dag at forholdet er det motsatte.

En del av de statistikker som utarbeides er særlig viktig å få bearbeidd så raskt som mulig. Skal vi bearbeide manuelt et materiale som er likt fra år til år, må vi gjennom den samme bearbeidingsrutinen hvert år. Og bearbeidingen kan vanligvis først begynne når materialet er helt ferdig revidert. Har vi lagt opp til maskinell bearbeiding, vil vi kunne bruke de samme EDB-programmer fra år til år. Bearbeidingstiden vil da normalt bli lavere ved andre gangs og seinere gangers "gjennomkjøring" enn første gang. Dette bør vi også vurdere når vi skal ta stilling til valg av bearbeidingsmetode: Om materialet bør bearbeides hurtig, og om det vil bli behov for den samme bearbeidingen ved seinere anledninger.

En del beregninger på grunnlag av det reviderte statistiske materiale vil i praksis være umulig å få gjennomført uten å ta EDB-maskiner til hjelp. Dette gjelder spesielt kompliserte matematisk-statistiske beregninger, men også krysstabuleringer i nevneverdig omfang. Ved hjelp av en EDB-maskin er det raskt gjort å få sortert om materialet, slik at vi kan få listet ut materialet i tabeller gruppert etter verdier på kjennemerker som vi er interessert i. En krysstabulering vil vi ha når vi ønsker å dele opp materialet etter flere forskjellige kjennemerker, og hvor de samme enhetene har kjennemerkeverdier som kan plasseres ikke bare i samme grupper, men i forskjellige grupper, altså må datamaterialet sorteres om før tabeller kan lages.

2. Manuell bearbeiding

Etter hvert som maskinell bearbeiding blir stadig billigere (pga. lavere priser på EDB-maskiner) og manuell bearbeiding dyrere (pga. at arbeidslønningene stadig stiger, og arbeidskraft er den viktigste kostnadskomponenten i manuell bearbeiding) skjer det en forskyvning over til at stadig mer av statistiske undersøkelser bearbeides maskinelt. Fremdeles brukes likevel manuelle teknikker i en viss utstrekning, spesielt i tilknytning til maskinell bearbeiding.

I en mellomstilling mellom maskinell og hva vi betrakter som manuell bearbeiding står bearbeiding ved hjelp av ulike former for hullkortmaskiner: sorterere, tabulatorer, collatorer og kalkulatorer. Slike maskiner er nå praktisk talt ikke i bruk.

Manuelle regnemaskiner, enten de er mekaniske eller elektroniske, er i dag stort sett hel-automatiske. På en helautomatisk maskin behøver man bare "slå inn" på en tangent den operasjon en vil maskinen skal utføre, og så gjør den det. Noen maskiner utfører bare addisjoner og subtraksjoner, mens andre også utfører divisjoner og multiplikasjoner. I de siste årene har det kommet små elektroniske bordregnemaskiner som, i tillegg til å utføre de mest vanlige regneartene, kan utføre visse mer kompliserte beregninger, f.eks. direkte prosentregning, potensoppløfting og kvadratrotutregning, beregning av gjennomsnitt og standardavvik. Disse maskinene har da innebygd programmerte instruksjoner for hva de skal gjøre hvis f.eks. gjennomsnittstasten trykkes inn. Det siste på markedet er små programmerbare bordregnemaskiner.

Bordregnemaskinene er både skrivende og ikke-skrivende; dette gjelder så vel de mekaniske som de elektroniske.

Når man skal bruke en manuell regnemaskin, lønner det seg alltid på forhånd å sette seg inn i hvordan maskinen virker. Dessuten må regneproblemet tenkes igjennom før man starter selve beregningsarbeidet.

3. Koding og dataoverføring

Når dataene er ferdig revidert, må de bearbeides videre før resultater kan offentliggjøres (publiseres). Bearbeidingen, som foregår etter revisjonen, kan vi dele i tre deler: 1) Koding og dataoverføring, 2) Gruppering og opptelling og 3) Utregning av statistiske mål. Ofte vil koding foregå samtidig med revisjon. Alle tre delene vil bli nærmere behandlet seinere. Vi skal først gi en oversikt sett ut fra et planleggingssynspunkt.

Skal data bearbeides maskinelt, vil en dataoverføring si det samme som at dataene overføres fra skjemaer og til en maskinlesbar form, f.eks. hullkort eller magnetbånd. En EDB-maskin må, for å utføre de operasjoner vi ønsker utført (gruppering og opptelling og utregning av statistiske mål), ha dataene overført til en form som den er i stand til å gjøre noe med, den må ha kodet dataene på en bestemt måte. Kodingen vil til en viss grad avhenge av hvilken type EDB-maskin vi skal operere med fordi enkelte maskiner med tilleggsutstyr kan gjøre ting som andre maskiner ikke kan gjøre. Noen EDB-maskiner har utstyr som gjør dem i stand til å lese dataene direkte (optisk) fra skjemaer (postgiroblanketter f.eks.), mens andre maskiner ikke har dette tilleggsutstyret, og dataene må da punches først (f.eks. på hullkort) og så eventuelt overføres til magnetbånd eller andre lagringsmedier.

En manuell bearbeiding vil si at det er personer som grupperer, teller og regner ut statistiske mål. Maskinell bearbeiding vil si at maskinen ved hjelp av programmerte instruksjoner utfører de samme (eventuelt andre) operasjoner vi ønsker utført.

En manuell bearbeiding tar ofte sitt utgangspunkt direkte i skjemaene, eventuelt etter at skjemaene er kodet. Koding vil si at man omgjør alle ikke-kvantitative kjennemerkeverdier til kvantitative verdier, dvs. siffer. Ofte vil skjemaene helt eller delvis være prekodet, dvs. at de forskjellige svaralternativer er gitt koder, m.a.o. at skjemautfylleren selv koder sine svar. På enkelte skjema er det avsatt plass til koding av svarene. Når en ved konstruksjon av skjemaer skal velge mellom åpent og bundet svaralternativ, vil det i første rekke være avhengig av om en tror alternativene er forståelige og greie for den spurte og bør i annen rekke være motivert med sparing av koding etc.

Av og til kan det være en fordel å overføre alle svar på et skjema til en spesiell kodingsblankett (kodeark). Denne kodingsblanketten vil da bestå av to og to kolonner: den ene er for spørsmålsnummereringen og den andre for den kodede verdien.

På en del blanketter er det heller ikke tilstrekkelig nummerering av spørsmålene, eventuelt at de er alfanumeriske. Da må hvert enkelt spørsmål gis hvert sitt nummer under kodingen.

Når man skal kode et datamateriale, vil bearbeidingsmåten ha betydning for kodingen fordi det er flere metoder for koding, dvs. metoder for å omforme ikke-kvantitative verdier til kvantitative verdier. Kodingen må fastlegges på forhånd, slik at hele materialet blir kodet ensartet. Uttømmende kodelister må derfor settes opp på forhånd. Ofte kan standarden (f.eks. Standard for næringsgruppering) brukes som kodeliste. En bør forsøke å gjøre bruk av standardiserte koder fordi det er en fordel ved eventuelle samkjøringer, integreringer, av to forskjellige materialer at en har kodet likt. Har kjennemerkeverdien "ugift" blitt gitt verdien 1 i et materiale, vil det være en fordel om "ugift" gis verdien 1 også i det andre materialet hvis de to verdiene skal stilles opp mot hverandre (f.eks. forandring i ekteskapelig status over en gitt periode).

Den viktigste måten å få overført dataene til EDB-maskin, har vært punching, og slik vil det fortsatt være en stund. Dataoverføring brukes også som betegnelse på det å overføre dataene fra hullkort til magnetbånd, dvs. fra en maskinlesbar form til en annen. Grunnen til at dataene oftest overføres fra kort til bånd er at data på bånd er lettere å 1) bearbeide (det går raskere å sortere om dataene, lese gjennom dataene etc. når de er lagret på bånd enn på kort), 2) lagre for seinere bruk (på ett magnetbånd kan man lagre like mange data som i kassevis med hullkort) og 3) kombinere med andre data (jfr. det arkivstatistiske system).

Tidligere ble data utelukkende punchet på hullkort, nå er det blitt mer og mer vanlig å punche dataene inn på såkalte disketter, som er runde magnetbåndskiver. Prinsippet her er at istedenfor å lage et hull lager punchmaskinen magnetiske felt. I tillegg til at selve puncheoperasjonen er blitt mye mindre støyende, tar diskettene mye mindre plass (en diskette som er like stor som en "single" grammofonplate, rommer like mange punchetegn som en hel eske hullkort), og de kan brukes flere ganger. De kan også sendes billig i posten. Det er dessuten mulig å programmere de nye punchmaskinene, dvs. at visse kontroller kan utføres under punchingen.

Noen data som vi ønsker å lagre er av en slik art at vi må ha dem raskt og ofte (f.eks. personnummer, adresse e.l.). Da vil det ofte være en stor fordel å ha lagret disse dataene, slik at de kan nås direkte (direkte aksess) ute noen blaug i lister, kort e.l. Slike data overføres derfor ofte til spesielle databærere som plater, trommer o.l., hvor data om den enkelte enhet kan tas fram direkte (jfr. begrepet databank). I statistikkproduksjon vil en som regel ikke ha behov for opplysninger om den enkelte enhet, derfor kan dataene overføres og lagres på databærere med sekvensiell adgang (magnetbånd og hullbånd).

4. Maskinell revisjon med oppretting

Gangen i en maskinell revisjon

Når et datamateriale skal revideres maskinelt, er det som oftest gjenstand for følgende operasjoner: Først overføres materialet til en maskinlesbar form, deretter kontrolleres materialet ved hjelp av et EDB-kontrollprogram, og som resultat får vi ut en liste som viser hvilke feil det ifølge kontrollen er i materialet. (En del av kontrollene kan være fastlagt slik at det ikke nødvendigvis er feil alt det som blir utskrevet, men bare data en ønsker nærmere vurdert.) Deretter rettes de gale dataene i materialet ved at nye riktige data overføres til en maskinlesbart form (f.eks. ved å punches), og ved hjelp av et EDB-opprettingsprogram byttes de gale dataene ut med de opprettede. Når opprettingen er ferdig, "kjøres" materialet på nytt gjennom kontrollprogrammet. Er det fremdeles noen feil, eller har vi rettet galt, må vi på nytt gjennom opprettingen. Slik kan vi fortsette helt til vi får et datamateriale vi godtar.

Vi har foran skissert arbeidsgangen når vi har et kontrollprogram. Ofte kan det være flere kontrollprogrammer som et datamateriale går gjennom. Samtidig vil vanligvis et kontrollprogram inneholde flere enn én kontroll.

På figur 9 på neste side (Maskinell revisjon med oppretting), er tegnet inn en vanlig arbeidsgang når vi har to kontrollprogrammer. Det kan godt tenkes at et materiale går gjennom flere kontrollprogrammer. De opprettinger som må foretas etter at vi har fått feillister, kan være helt enkle (f.eks. kan det "bare" være feilpunchinger), eller vi må kanskje foreta fornyet manuell revisjon av visse enheter.

På figuren er tegnet inn at to forskjellige opprettingsprogrammer er brukt. Det kan godt hende at ett og samme opprettingsprogram kan brukes begge steder (hvis f.eks. opprettingsprogrammet går ut på å bytte ut et tall med et annet).

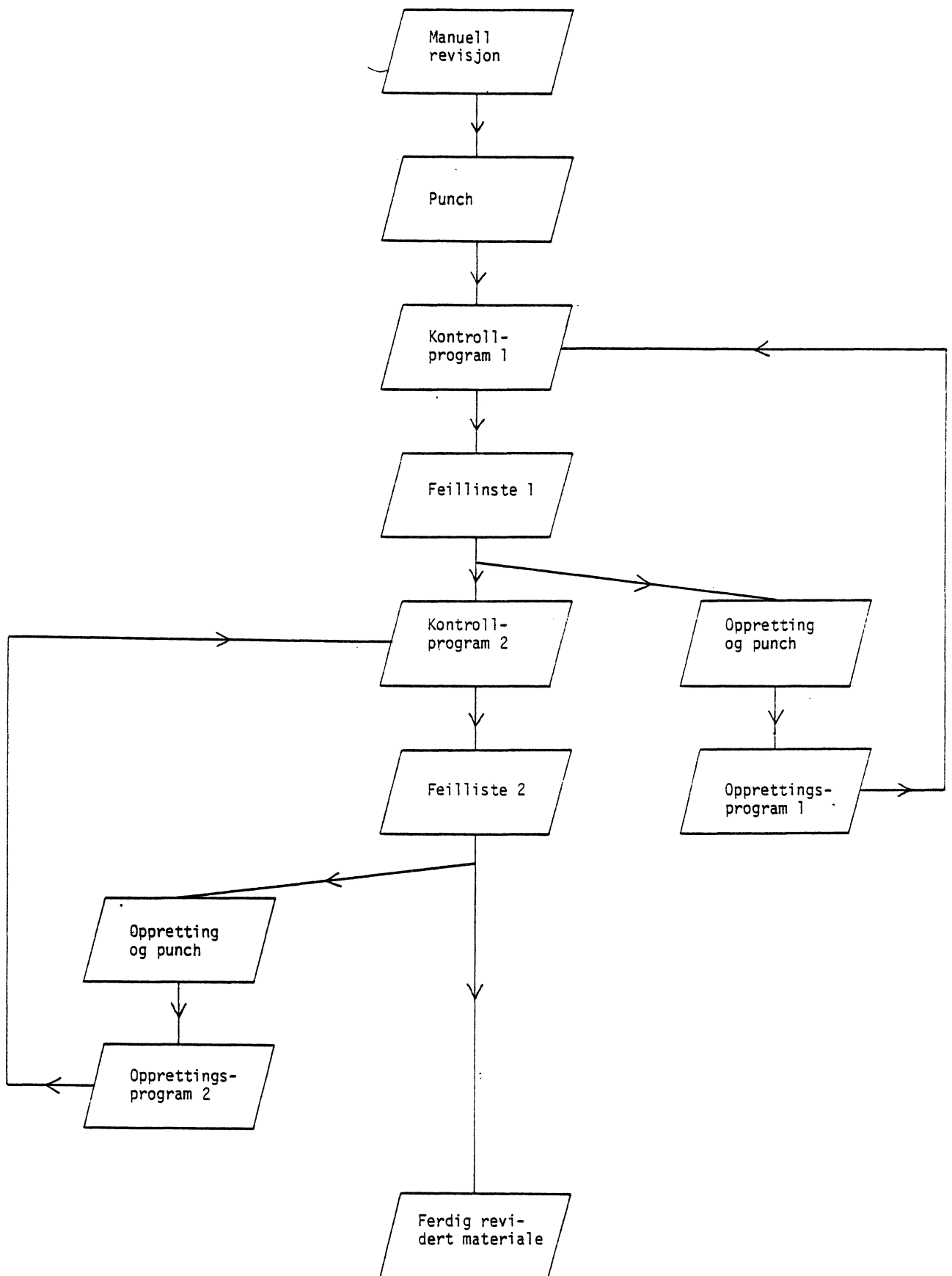
Tidligere har vi også vært inne på muligheten av en viss manuell revisjon samtidig med eller etter den maskinelle revisjonen. Dette er ikke tegnet inn på figuren. Vi kan likevel tenke oss en eventuell slik manuell revisjon plassert sammen med opprettingen som er tegnet inn, eller som et nytt parallellogram rett foran "Ferdig revidert materiale".

De enkelte kontrollprogrammene kan godt tenkes å inneholde flere kontroller enn bare en, men hvert slikt program pleier bare å foreta en type kontroller. F.eks. kontrollerer program 1 om alle de punched kortene er punchet riktig (f.eks. om det er riktige tegn på riktige steder, se seinere), mens program 2 kontrollerer om tallene er riktige (f.eks. sumkontroller). Det er en fordel ikke å foreta for mange kontroller om gangen hvis det fører til store feilutskrifter.

Kontrollprogrammene kan som nevnt også inneholde de samme kontroller som tidligere er foretatt manuelt. Vi foretar likevel ofte dobbeltkontroller enten for å se om alle revisorer har revidert likt både seg imellom og over tiden, eller fordi den manuelle revisjonen er lagt opp nettopp med sikte på en maskinell etterkontroll.

En del av de revisjonskontroller vi ønsker å gjennomføre er av en slik art at vi må trekke noen opplysninger (data) ut fra hvert skjema og så se på alle disse dataene under ett for å vurdere om noen er tvilsomme eller gale. Skal dette gjøres manuelt, koster det mye tid og dermed penger. Derfor kan det hende slike kontroller ikke gjøres. Maskinelt er det lett å foreta omgrupperinger og sorteringer av materialet, og når det er nødvendig med en eller flere sorteringer av materialet (skjemaene) før det kan revideres, vil det ofte være en fordel om revisjonen kan utføres maskinelt.

Figur 9. Maskinell revisjon med oppretting



Kort oversikt over kontrollmetoder i Statistisk Sentralbyrå¹⁾

De vanlige kontrollmetoder som Statistisk Sentralbyrå bruker, er:

- (1) Gyldighetskontroller for koder (f.eks. at en næringskode er mulig)
- (2) Kontroll av mulige kombinasjoner av koder eller kombinasjoner av koder og størrelser (f.eks. at et barn ikke er gift)
- (3) Kontroll av at forskjellige typer av funksjoner mellom de kjennemerkene som skal kontrolleres, ligger mellom visse grenser. Et spesialtilfelle er her kontroll av rent bokholderitekniske definisjonssammenhenger. Ved siden av definisjonssammenhenger bruker en i Statistisk Sentralbyrå i stor utstrekning kontroll av enkle forholdstall mellom to størrelser (f.eks. lønn pr. sysselsatt)

I Byrådet har en også latt både størrelser for den enkelte enhet fra den periode som skal kontrolleres og foregående periode gå inn i kontrollfunksjonene. (En har f.eks. tillatt lønn pr. ansatt å stige maksimalt med 20 prosent fra foregående år for en og samme bedrift.) På mer aggregert nivå, dvs. for grupper av enheter, foretas det svært mange slike sammenlikninger med foregående periode. Selve fastleggelsen av kontrollmetodene foregår skjønsmessig på grunnlag av spesialistenes (revisorenes) erfaring mv. på de enkelte statistikkområder. Noen mer analytisk sammenlikning av fordeler og mangler ved de forskjellige kontrollmetoder har en ikke foretatt. Det er nå utviklet kontrollprogram som tar opptil 1 000 kontroller i en kjøring.

Automatisk spesifisering av kontrollkriterier kontra skjønsmessig spesifisering av kontrollkriterier

Med automatisk spesifisering mener en at kontrollkriteriene beregnes maskinelt etter en eller annen metode. Med skjønsmessig spesifisering mener en at de fastsettes rent skjønsmessig av spesialistene på vedkommende fagstatistikk. Denne sontringen gjelder da særlig kriteriene i kontrollfunksjonene. Ved kombinasjonskontrollene vil det i stor utstrekning være rent logiske sammenhenger som fagspesialistene på vedkommende statistikkområde har greie på.

De moderne EDB-maskiner skulle gi store muligheter for automatiske beregninger. Statistisk Sentralbyrå har likevel foreløpig ikke gått så særlig langt i automatisk spesifisering på dette området.

Ellers brukes i stor utstrekning skjønsmessig spesifisering av kontrollkriteriene. For engangstellingene foretas nå og da gjennomkjøringer av et utvalg av materialet for manuell inspeksjon av bruk av et sett av kontrollkriteriene. Ved priskontrollen på den månedlige konsumprisindeksen kjører en ut fordelingene av prisene for hver enkelt vare. Prisgrensene fra foregående måned kontrolljusteres manuelt ved manuell inspeksjon av disse fordelingene.

Automatisk oppretting av feil

Statistisk Sentralbyrå har ikke kommet noen særlig vei her når det gjelder automatisk oppretting av feil. Det eneste unntaket er en kontroll med den månedlige konsumprisindeksen. En retter her opp feil ved å bruke den gjennomsnittlige prisendring for samme vare innen samme marked (et marked er f.eks. bygdene innen en landsdel). En listekjører så de automatiske opprettingene, og spesialistene på vedkommende statistikkområde kontrollerer manuelt de automatiske opprettingene. Ved bedriftstellingen 1974 ble manglende data i visse tilfelle beregnet automatisk.

Ellers er en vel nå noe mer betenkt med hensyn til automatiske opprettinger enn en var for noen år siden. Det som her kommer inn og som de nye EDB-maskinene har muliggjort, er:

- (1) Opprettelsen av dataarkiver med utkjøring av statistikk på tvers av statistikkområder og på tvers av tiden
- (2) Mer analytisk utnyttelse av enkeltdataene (f.eks. regresjonsberegninger)

Begge disse forhold gjør at det ikke er mulig å forutsi hvilke kombinasjoner av dataene som en vil søke å utnytte. Det stilles derfor større krav til kvaliteten av dataene.

1) Dette avsnittet og de etterfølgende avsnitt c-f om maskinell kontroll er i hovedsak hentet fra et foredrag Nils Bakke, Statistisk Sentralbyrå, holdt på et seminar på Kongsberg 22.-25. mars 1971. Avsnittene er noe endret og forkortet.

Forsøk på vurdering av hvor langt en skal gå i automatisering

I hullkortmaskinenes første tid var det en sterk strid om automatisk og maskinell kontra manuell kontroll av materiale. Denne striden er slutt, og det er alminnelig godkjent at automatisk kontroll av materiale byr på store fordeler.

Den viktigste grunnen til dette er antakelig vanskeligheter med en nøyaktig fastlegging av kontrollkriterier ved automatisk spesifisering. Ved første gangs spesifisering av kontrollkriteriene trenger en et forholdsvis stort materiale som er sammenliknbart med det materiale som skal kontrolleres. Ved årsstatistikk kan det f.eks. være nødvendig å gå flere årganger tilbake for å få tilstrekkelig antall enheter. Men automatisk bestemte kontrollkriterier vil da hurtig bli foreldet og ubrukelige. Antall enheter for hver ny årgang vil ofte også være for lite til noen god ajourføring av kontrollkriteriene. Til sist kan nevnes at selve gjennomkjøringen av et tilstrekkelig stort materiale for første gangs spesifisering av kontrollkriteriene vil være en kostbar investering.

Det følger av dette at fullstendig automatiserte systemer særlig er aktuelle for månedsstatistikker og meget store datamasser (som f.eks. en telling).

En kan også anføre argumenter for at det ikke er klokt av hensyn til den "menneskelige faktor" å presse automatiseringen for langt. Det personale på et fagkontor som skal utføre selve kontrollarbeidet, har som regel liten teoretisk bakgrunn og av den grunn kanskje mindre forståelse for mer innviklete beregninger. Dette fører for det første til at de ikke har noe særlig å gi når det gjelder selve utviklingen av det nye systemet. Systemet vil dermed ikke bli deres eget. De vil også ha vanskeligheter med å praktisere de manuelle inngrep i systemet som antakelig vil være nødvendig.

Muligheter for integrering av kontrollprosessen ved en statistisk bearbeiding

Det har lenge vært til stor ergrelse og bekymring i de statistiske kontorer at revisjonen tar så lang tid. Særlig de forskjellige stadiene med fram og tilbake mellom fagkontorer (statistikkbearbeidende kontorer) og EDB-avdeling tar lang tid. Overgangen fra hullkortmaskiner til EDB-maskiner medførte en viss integrering av selve den maskinelle prosessen. Men utover dette har det ikke skjedd noe framskritt på dette området.

Uten tvil vil de nye dataskjermer med tastatur medføre en liten revolusjon. Med denne teknikken kan vi utføre en rekke maskinelle kontroller under selve dataoverføringen. Selve dataoverføringen kan også gjøres direkte som en del av revisjonen. En får altså integrert dataoverføring og revisjon. Det er store muligheter for tidsbesparinger og kostnadsreduksjoner. Dessuten synes det å bli stadig vanskeligere å få folk til bare å sitte med monotont dataoverføringsarbeid. Foreløpig er slikt skjermutstyr og integrering av revisjons- og dataoverføringsfunksjon bare på prøvestadiet i Byrået, men erfaringene hittil er meget positive.

En må også være klar over at lang revisjonstid innebærer større fleksibilitet mht. svarfrist. Svarprosenten øker vanligvis ved romslig oppgavefrist.

Generelle kontrollprogrammer

Opplistingen av kontrollmetoder (side 64) forteller at det her må være felles trekk som går igjen på de forskjellige statistikkområder. Tidligere programmerte en disse kontrollmetodene på ny for hver statistikkbearbeiding.

Statistisk Sentralbyrå har arbeidd en del med å utvikle generelle kontrollprogrammer (f.eks. KOKS) og har nå i drift et generelt kontrollprogram som tar gyldighetskontroller og kombinasjonskontroller (jfr. side 64). I det generelle kontrollprogrammet er det mulig å foreta addisjoner, multiplikasjoner og divisjoner av størrelser og å kontrollere om resultatstørrelsen ligger mellom visse grenser (jfr. side 64).

B. Bearbeidings- og resultatkontroll

I denne fasen beregner vi hvor god undersøkelsen har vært i den forstand at vi ser etter om vi har fått med alle enhetene i massen, eventuelt hvor stor dekningsen er (svarprosent eller dekningsprosent). Videre er vi interessert i å få rede på om de enhetene som har besvart våre skjemaer er en skjev delmasse eller ikke. Har vi f.eks. sendt skjemaer til like mange menn som kvinner, men fått svar fra dobbelt så mange menn som kvinner, vil de innsendte oppgavene kunne gi et skjevt bilde av det vi er ute etter å belyse.

Vi grupperer derfor de innsendte enhetene for å se om svarprosenten avviker fra gruppe til gruppe. Selv om begrepet representativ undersøkelse går på selve utvalgsmetoden (dvs. hvordan enhetene vi har spurt er kommet med i utvalget), vil forskjeller i svarprosentene for enkelte grupper kunne ødelegge utsagnskraften til de konklusjoner vi måtte ønske å trekke på grunnlag av en undersøkelse. Ødelagt eller redusert utsagnskraft vil si at eventuelle konklusjoner vil bli mer usikre.

Det er på dette trinn av en undersøkelse vi bør vurdere en del av de totaltallene vi kommer fram til for å se om de er sannsynlige. Har f.eks. alle vi har spurt artium eller høyere utdanning, mens det etter utvalgsplanen bare er ca. 30 prosent som skulle ha denne utdanningen, er noe galt. Det kan skyldes spørsmålsformuleringen, revisjonen, kodingen eller dataoverføringen. Dette må undersøkes, og det må gjøres før vi begynner å beregne statistiske mål som skal belyse vår problemstilling. Det er fire typer av feil som kan gjøres under bearbeidningen: 1) Grupperingsfeil, 2) Tellingsfeil (feil ved selve undersøkelsen), 3) Overføringsfeil og 4) Revisjonsfeil. Vi må på forhånd under planleggingen fastlegge kontrollmulighetene av feilbearbeiding. Ved kontroller må vi ikke bare finne at det er begått feil, men finne også hvor og hvorfor det er begått feil for å avdekke svake ledd, enten det nå skyldes opplegg eller revisorer.

Seinere vil vi gå nærmere inn på ulike måter å stille sammen de innsamlede data på. Dataene må stilles sammen slik at de karakteriserer undersøkelsens masse. Vi må da se om de statistiske mål vi kommer fram til faktisk belyser den problemstillingen vi hadde som formål å belyse.

Det er mange ulike statistiske mål: summer, gjennomsnitt, hyppigheter, spredninger, fordelinger, regresjoner osv. Det bestemte mål vi vil velge, vil avhenge av problemet vi ønsker belyst. Er det f.eks. et årsaksforhold vi vil undersøke, vil vi være interessert i å stille de faktorene som vi på forhånd tror er forklaring opp imot resultatet, slik at våre antakelser kan konfronteres med de observerte data.

En del statistiske mål medfører kompliserte og store beregninger, og de vil da med fordel kunne utføres maskinelt. Valget mellom maskinell og manuell bearbeiding vil ofte avgjøres på grunnlag av hvilke statistiske mål vi ønsker å få beregnet.

Det er ved hjelp av statistiske mål vi får komprimert ned dataene vi har samlet inn, til et nivå som gir oss en oversikt over massen og enhetene vi ønsker å si noe om. Vårt valg av slike mål må derfor være veloverveide, f.eks. vil vi ikke si det er behagelig å vaske seg under en spring med vann som har gjennomsnittstemperatur 25°C hvis vannet kommer støtvis iskaldt og kokende. Et slikt problem måtte derfor beskrives med f.eks. et mål for jamnhet i temperaturen i tillegg til gjennomsnittstemperaturen.

Velger vi under planleggingen at bearbeidningen skal foregå maskinelt, kan vi seinest når skjemaene er ferdig utarbeidd, begynne med programmeringen (dvs. utarbeidelsen av de maskinelle rutinene). Programmene vi har bruk for kan da være ferdige til vi skal begynne med selve bearbeidningen.

Det er ofte ved hjelp av statistiske mål, f.eks. gjennomsnitt, at vi kan vurdere om resultatene vi har kommet fram til er rimelige. M.a.o. er det ofte at vi først foretar de endelige kontrollene når materialet er ferdig bearbeidd. Sluttkontrollen av en statistikk foretas som regel på grunnlag av manuskriptet til publikasjonen.

C. Standardisering

Svarene på åpne spørsmål skal under bearbeidingen ofte plasseres i grupper, f.eks. aldersgruppe 20-25 år, og disse gruppene - eller båsene - bør være definert og avgrenset på forhånd, dvs. før dataene hentes inn. I de tidligere omtalte standarder er det laget en hel rekke standardiserte gruppeinndelinger for kjennemerker vi innhenter verdier for.

Vi må definere de egenskapene vi skal registrere ved en undersøkelse, og gruppe- eller klasseinndelingen må fastlegges allerede i planleggingsfasen. Ved standardiseringen må vi ta stilling til de klasser som vi skal arbeide med seinere i undersøkelsen.

Når vi tar sikte på å gruppere svarene på åpne spørsmål i klasser, bør vi huske på det ved konstruksjonen av skjemaet, slik at kodingen av de åpne spørsmål kan gå greit.

6. ELEMENTÆR ANALYSE

Data som samles inn ved en statistisk undersøkelse representerer informasjon. Denne informasjonen ønsker statistikkprodusenten å formidle videre til brukerne. Produsenten bør prøve å komme fram til noen få tall som presenterer den innsamlede informasjonen på en oversiktlig måte. Ofte presenteres denne informasjonen som tall i tabeller, eller vi framstiller tallene grafisk i figurer (diagrammer og tematiske kart).

I boken Hva er statistikk? av Wallis & Roberts er det gitt mange eksempler på representasjon av data. I kapittel 7¹⁾ i boken har forfatterne tatt for seg sammenhengen mellom observasjon og måling, dvs. mellom den virkelige verden og de tallene en statistiker opererer med. Ofte stilles det krav til målingen, og kravene uttrykkes i samfunnsvitenskapene som regel ved en av følgende tre begreper:

- Formålsdekning (validitet). At en undersøkelse (tallfesting, måling) har høy formålsdekning, betyr at undersøkelsen nettopp har målt det den har til formål eller hensikt å måle
- Pålitelighet (reliabilitet). En undersøkelse har høy pålitelighet når den bygger på et godt register, når spørsmålene er presist formulert og når alle ledd av innsamlingen, besvarelsene og videre bearbeiding er under behørlig kontroll
- Presisjon. En undersøkelse har høy presisjon når nøyaktigheten er høy. Presisjon og nøyaktighet er altså synonymer, og betegnelsen presisjon brukes oftest i forbindelse med utvalgstellinger og da som uttrykk for hvor godt vi ved hjelp av et utvalg får målt egenskapene slik de faktisk er i hele undersøkelsesbestanden. Vi kan altså øke presisjonen ved å øke utvalget

Tallene statistikeren opererer med er enten direkte målinger, observasjoner, registreringer av hendinger etc. eller en sammentrenging av disse målingene i statistiske mål eller representanttall. Slike representanttall er f.eks. summer, gjennomsnitt, fraktiler (bl.a. median) osv. Kvaliteten (jfr. pkt. a-c ovenfor) på representanttallene kan aldri bli bedre enn kvaliteten på de tallene som er grunnlaget for beregningene.

Den viktigste og mest brukte måten å få fram egenskaper ved kjennemerkene (fordelingsegenskaper) er tabelloppstillingen. En tabell er en systematisk og komprimert ikke-verbal oppstilling av et statistisk materiale. Som regel inneholder en tabell bare gruppeopplysninger og ikke individualopplysninger, dvs. opplysninger om enkeltenheter.

1) Kapitlet i boken bør absolutt leses.

A. Gruppering og opptelling

1. Tabeller

En tabell består av en (tabell-)overskrift, et (tabell-)hode, en forspalte, kolonner og linjer med tall, eventuell kildehenvisning og fotnoter (forklaringer mv.). Vi opererer sjelden med tabeller med mer enn tredobbelt hode, isteden deler vi tabellen i to.

Figur 10

Forspalte- hode	Tabellhode (enkelt)

lin-
jer

Figur 11

	Tabellhode (tredobbelt)			

kolonner

En tabell med enkelt hode bruker vi når vi skal fordele en masse av enheter etter ett kjennemerke (en éndimensjonal absolutt fordeling). Skal massen fordeles etter to kjennemerker, bruker vi en tabell med dobbelt hode (todimensjonal fordeling). Istedenfor å ha to kjennemerker i tabellhodet, kan vi plassere det ene kjennemerket i forspalten. Begge kjennemerkene kan da anta flere alternative verdier, og disse verdiene er gitt i hodet og forspalten (f.eks. næring og år).

I én og samme tabell kan det av og til være ønskelig å tabellere to eller flere masser i samme tabell, f.eks. i en næringsoversikt gi både tall fra produksjon og antall sysselsatte.

Ofte er verdiene på kjennemerkene kvantitative, og kjennemerker kan ofte også anta uendelig mange verdier, derfor må en dele inn kjennemerkeverdiene i klasser.

Tabellen nedenfor er en generell toveis tabell (todimensjonal tabell med tabellhode og forspalte), hvor vi antar at det er to kjennemerker hos enhetene i den undersøkte massen vi er interessert i: kjennemerke I og kjennemerke II. Begge kjennemerkene kan anta et uendelig sett av verdier. Vi deler her imidlertid disse verdiene inn i klasser: Verdiene som kjennemerke I kan anta er enten i klasse 1, klasse 2, klasse 3 osv. opp til et bestemt tall, f.eks. klasse 6. Kjennemerke II kan anta verdier som vi plasserer i klasse a, klasse b, klasse c osv., helt opp til en bestemt bokstav, f.eks. e.

Enhetene vi har undersøkt kan vi altså enten plassere i klasse 1, 2, 3 osv. Enten er det antall enheter med f.eks. kjennemerkeverdi 2 som plasseres i ruten for klasse 2 eller en sum av kjennemerkeverdiene for enhetene som har "klasse 2-verdien". Tallet i ruten for klasse 2 gir da uttrykk for hvor ofte, hvor hyppig kjennemerkeverdien "klasse 2" forekommer, altså en hyppighet.

Figur 12

		Kjennetegn I						Sum
		1	2	3	4	5	6	
Kjenne- tegn II	a	n_{a1}	n_{a2}	n_{a3}	n_{a4}	...		Marginal for- deling etter kjennetegn II
	b	n_{b1}	n_{b2}	...				
	c	.						
	d	.						
	e	.						
Sum		Marginal fordeling etter kjennetegn I						

I denne generelle tabellen har vi tegnet inn en sumkolonne og en sumlinje. Alle enhetene i massen som har en kjennemerkeverdi for kjennemerke II som ligger innenfor grensene til klasse a, har vi gruppert i den første linjen. I tabellen står det f.eks. n_{b2} , det vil si antall enheter som har både kjennemerkeverdi b (dvs. innenfor klassen b) og verdi 2. Alle enhetene som har verdien b for kjennemerke II, er det vi summerer opp i sumkolonnen. I sumkolonnen har vi altså den marginale fordelingen etter kjennemerke II. Det er den marginale fordelingen fordi vi her har fordelt enhetene bare etter verdier på kjennemerke II. Tilsvarende er sumlinjen den marginale fordelingen etter kjennemerke I.

Hver eneste tabell bør kunne forstås uten tekstkommentarer. Tabellen bør derfor være selvfor-klarende, slik at leseren ikke behøver lete i en tilhørende tekst etter forklaringer. Ord som brukes i tabellen må være forståelige for den som bare leser tabellen og ikke eventuelle tilhørende tekster, men selvfølgelig skal ikke en tabell inneholde for mye definisjoner og forklarende tekst heller, ellers vil den lett bli uoversiktlig og vanskelig å lese. Fotnoter (noter) og annen forklarende tekst er ofte nødvendig, men bør så vidt mulig unngås. Fotnoter, forklaringer, kilder plasseres under tabellen.

Bygger en tabell på primærstatistikk, vil kilden for tabellen i sin alminnelighet være klar, men hvis en tabell er utarbeidd på grunnlag av andre tabeller eller på andre måter er en sekundærstatistisk presentasjon, må kilden(e) oppgis i tabellen.

Over hver kolonne i tabellen må angis hvilken måleenhet kolonnens tall er i (personer, kroner, stykker, tonn osv.). Har alle kolonner samme måleenhet, plasseres enhetsangivelsen i tabelloverskriften. Ofte vil man imidlertid i en tabell gi både de absolutte verdier for hvor mange enheter som bærer et kjennemerke og de relative verdiene. I den generelle tabellen er alle enhetene som har verdi 2 på kjennemerke I lik $n_{a2} + n_{b2} + n_{c2} + \dots = n_{.2}$. De tilsvarende relative hyppighetene er $n_{a2}/n_{.2}$, $n_{b2}/n_{.2}$, $n_{c2}/n_{.2}$ osv. Ofte angis de relative hyppigheter i prosent; da må tallene multipliseres med 100, og f.eks. vil $100 \cdot n_{b2}/n_{.2}$ si hvor mange prosent av alle enhetene med verdien 2 på kjennemerke I som har verdien b på kjennemerke II. Vi kan gi både absolutte og relative tall i én og samme tabell. I tabellen bør de relative tallene plasseres i kolonnen ved siden av de absolutte tallene. Ofte vil det imidlertid være en fordel for oversiktens skyld å lage én tabell for de absolutte tallene og én tabell for de relative tallene.

I den generelle tabellen vi har tegnet her, er summene plassert lengst til høyre og nederst, men siden summene (de marginale fordelingene) er noe av det viktigste i en tabell, er disse ofte plassert på den øverste linje og kolonnen lengst til venstre, fordi man lettest ser tallene her. (Dette er også regelen i Byrået.)

Er datamaterialet inndelt i klasser, bør klassegrensene angis tydelig. Man bør derfor ikke i en tabell over antall bedrifter fordelt etter sysselsetting skrive grensene slik:

20 - 50	sysselsatte	
50 - 100	"	
100 - 200	"	osv.

men isteden

20 - 49	sysselsatte	
50 - 99	"	
100 - 199	"	osv.

En tabell bør være lett å lese, dvs. det bør være en enkel sak å følge en bestemt kolonne og en bestemt linje. Har vi en tabell med mange kolonner, bør disse nummereres, såfremt ikke teksten øverst i kolonnen innebærer en nummerering, f.eks. en næringsgruppeinndeling. Det tilsvarende gjelder for linjene.

For at en tabell skal være lett å lese, vil uthevelse av de viktigste tallene være en fordel. Uthevelsen kan være i form av kursiv eller halv- og/eller helfete typer. De viktigste tallene er da summene (de marginale fordelingene) og eventuelle gjennomsnittstall. Det kan også være en fordel med en skillestrek mellom kolonner med tall som har ulik benevning, f.eks. kroner og prosenter. Man bør likevel unngå kolonnestreker og -linjer i en tabell hvis den skrives på maskin eller trykkes. Slike markeringer vil ofte gjøre det vanskeligere å lese en tabell, men brukt med varsomhet kan likevel oppstrekninger være nyttig for leseren.

Hvis man må dele en tabell på flere sider, bør tabelloverskriften, hode og forspalte gjentas. Fotnoter og eventuelle kildeangivelser står derimot bare på den første siden (henvisninger til denne siden gis når tilsvarende fotnotetegn forekommer på de andre sidene av samme tabell). Visse definisjoner og forklaringer kan plasseres i tabelloverskriften med litt mindre skrift enn selve navnet eller betegnelsen på tabellen. En slik mer nøyaktig forklaring på visse ting i tabellen, f.eks. på klassifiseringen av enhetene, kalles på fagspråket en ingress. Trolig er det bedre å plassere slike forklaringer mv. i én eller flere fotnoter.

En tabell bør være logisk oppbygd, eventuelle årsangivelser f.eks. bør være plassert slik at eldste år plasseres øverst eller til venstre og siste år nederst eller til høyre.

I Statistisk Sentralbyrå har man skriftlige regler for utstyr mv. i publikasjonene som utgis.¹⁾ Disse reglene sier ganske detaljert hvordan registre, tekst, figurer, tabeller, noter, kilder, litteraturlister etc. skal være utformet og redigert i Byråets publikasjoner. Reglene for tabeller sier f.eks. bl.a.: "Tidsperiode (f.eks. år) eller tidspunkt statistikken gjelder for skal normalt komme etter minste regionale enhet." Reglene sier også noe om ordbruk og skrivemåte. Videre er det gitt detaljerte regler for skrifttyper, grader mv., f.eks.: "På baksiden av tittelbladet skal overskriften settes med halvfet Bodoni i størrelsen 10 punkter." Hensikten med disse detaljerte reglene er å sikre at tabellene blir standardiserte, dvs. har man lest en tabell, vil neste tabell kunne leses lettere, osv. Dessuten vil man ved standardisering sikre at alle tabeller blir like bra, dvs. utarbeidd etter prinsipper som er gode og gjennomprøvede.

Den beste måten å lære å sette opp tabeller på er trolig å kopiere tabeller som er velkonstruerte. Tabellene i Statistisk årbok kan med fordel kopieres, selv om det også der finnes tabeller som rent teknisk kan bli bedre.

1) Regler for publikasjoners redigering og utstyr, nr. 40 i serien Statistisk Sentralbyrås Håndbøker.

Reglene tatt i bruk sommeren 1978. (Også tidligere var det skriftlige regler.)

Minst like viktig som å utarbeide en tabell riktig, er det å kunne lese en tabell riktig. I boken Hva er statistikk? av Wallis & Roberts er det i kapittel 9 ved hjelp av et eksempel gitt 10 regler for hvordan en tabell bør leses. De 10 reglene er:

1. Les overskriften grundig
2. Les ingressen eller andre forklaringer grundig
3. Legg merke til kilden. Er den pålitelig?
4. Se på fotnotene
5. Finn ut hvilke måleenheter som er brukt
6. Se på gjennomsnittet i totalen. (Sum av kjennemerkeverdiene dividert med antall enheter)
7. Se etter hvilke variasjoner det er i materialet
- 8 og 9.¹⁾ Se på de marginale fordelingene, og sammenlikn de enkelte kolonner og linjer med de marginale fordelingene. Her undersøker man om virkningene i totaltallene gjelder gjennomgående
10. Se etter avvik eller uregelmessigheter (anomalier)

Støtter man seg til disse 10 reglene når man går gjennom en tabell, vil man være i stand til å gi en kort oppsummering av den informasjon tabellen gir.

2. Grafisk framstilling²⁾

Enkle tabeller kan ofte med hell framstilles grafisk. Hvordan man i en publikasjon bør framstille visse resultater, formidle en informasjon, vil avhenge av både type av informasjon og bruker (mottaker). En kan ikke alltid forutsette at alle leserne kan lese en tabell. Dessuten vil en ved hjelp av en grafisk framstilling ofte lettere kunne poengtere viktige resultater, slik at tekstlige kommentarer til en tabell eventuelt kan gjøres overflødig. Siden vi har forskjellige hukommelsestyper (noen husker best det de ser, andre det de hører, leser osv.), vil en nå flere mennesker med informasjon ved å variere påvirkningsmidlene, f.eks. variere mellom tabeller og diagrammer.

Av og til brukes ordet figur som fellesbetegnelse for diagram og tematisk kart. Forskjellen mellom et diagram og et tematisk kart er at kartet i seg selv, uten hensyn til skraveringer og inntegnede diagrammer, gir en geografisk dimensjon til figuren. Et tematisk kart blir av og til også kalt kartogram.

Det finnes flere ulike typer av diagrammer³⁾:

1. Kurvediagram
2. Stolpediagram (inkl. histogram)
3. Flatediagram (inkl. sektordiagram)
4. Volumdiagram
5. Punktdiagram (spredningsdiagram)
6. Trekantdiagram
7. Fjærdiagram

Kombinasjoner av disse diagramtypene, spesielt 1 og 3, finnes også.

I boken Hva er statistikk? av Wallis & Roberts er det i kapittel 4 gitt noen eksempler på misvisende diagrammer. Disse eksemplene bør vi ha i minnet når vi nå kort skal se på de ulike typer av diagrammer. I publikasjonen Sosialt utsyn 1977 er gjengitt gode eksempler på de fleste av disse diagrammene.

1) Reglene 8 og 9 er her slått sammen til en regel og gitt en mer generell utforming enn i boken.

2) Kapitlet har trukket store veksler på et internt notat utarbeidd av Jan Byfuglien om Arbeidet med figurer til Statistisk Sentralbyrås publikasjoner. Dette arbeidet er kommet i serien Statistisk Sentralbyrås Håndbøker nr. 43.

3) Litt mer utførlig om dette i Praktisk statistikk av Erland v. Hofsten.

Diagram 1 kan også karakteriseres som et sammensatt diagram fordi en enkelt kurve (arbeidssøkere totalt) er delt opp i flere underkurver (over og under 25 år). (I dette diagrammet står det at tallene for de arbeidsløse for en gruppe er satt i forhold til arbeidsstyrken uten nærmere presisering av om det er hele arbeidsstyrken, eller bare gruppens arbeidsstyrke.)

Når vi skal tegne et slikt kurvediagram, er det noen regler vi bør følge:

- a) Antall kurver, bokstaver og tall i et diagram bør være så lite som mulig. Ved å variere strektypene (stiplet, prikket osv.) er det lettere å lese et diagram med flere kurver
- b) Forklaringen på de forskjellige kurvene i diagrammet må framgå tydelig av diagrammet eller av en medfølgende rubrikk
- c) Kildehenvisning må gis
- d) Skalaen må vies stor oppmerksomhet. Hvis vi ikke begynner på null, bør dette tydelig angis ved et brudd i diagrammet:

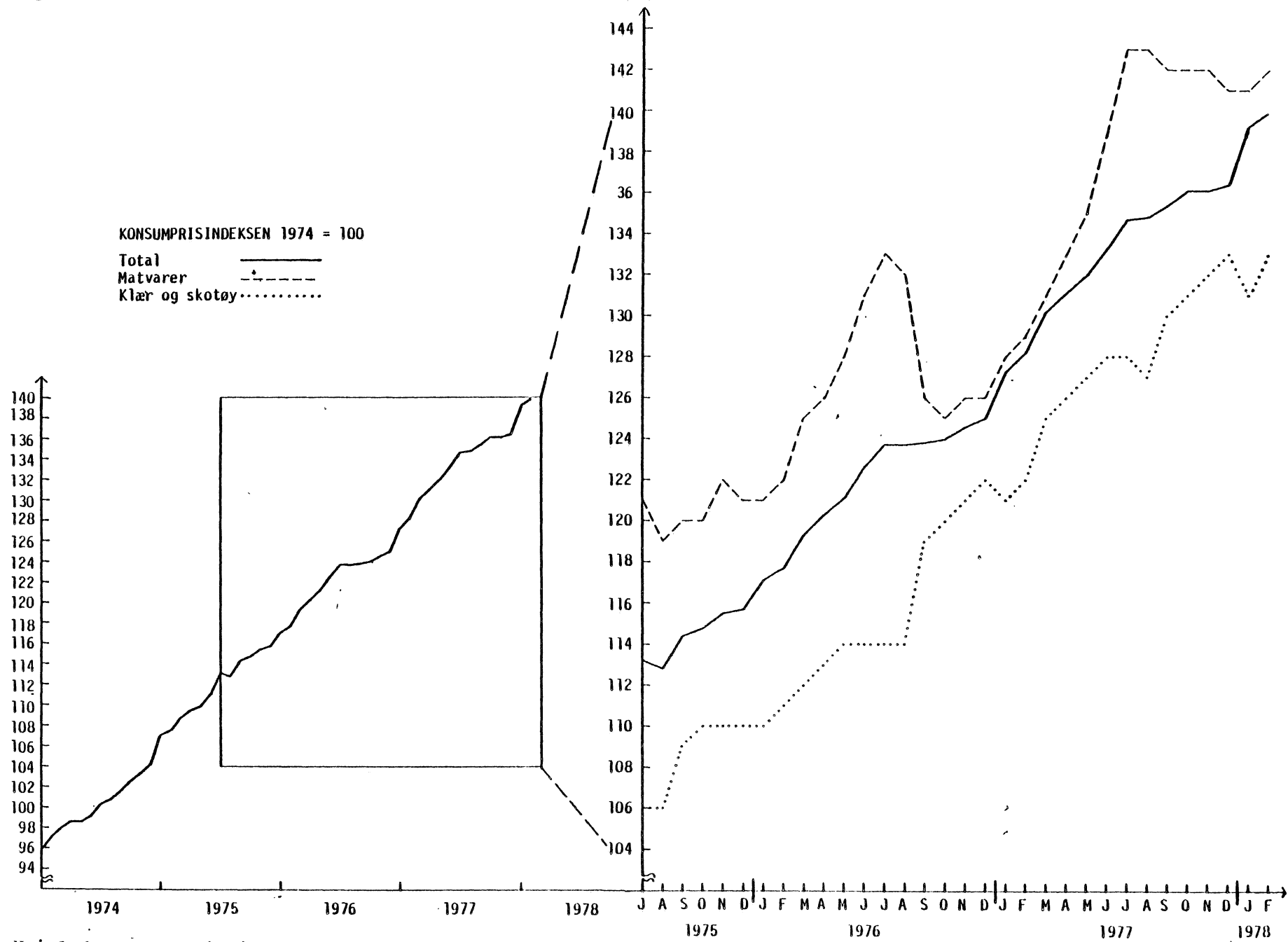


Ved indeksserier har man her et unntak, men her bør 100-linjen markeres særssilt.

- e) I et tidsdiagram bør vi angi om kjennemerkene refererer seg til et bestemt tidspunkt eller til en periode. Ofte bør en vurdere å bruke stolpediagram når dataene er gitt som gjennomsnitt for perioder
- f) For å få fram relative endringer gjør vi bruk av semilogaritmisk eller dobbellogaritmisk diagram. Ved semilogaritmisk diagram får vi fram den relative endringen i et kjennemerke, f.eks. ettersomtiden går. Ved et dobbellogaritmisk diagram får vi fram den relative endringen i et kjennemerke i forhold til den relative endringen i et annet kjennemerke

Hvis det er behov for en lang tidsserie, men hvor hovedvekten ønskes lagt på de siste årene, kan vi bruke et såkalt sammensatt diagram. Dette er et diagram hvor vi forstørrer de siste årene. Forstørringen går da enten ut på at vi for de siste årene også får fram variasjonene over det enkelte år (forstørrer den horisontale aksen), eller at vi for de siste årene tegner inn flere kurver (forstørrer den vertikale aksen), se figur 14. På denne figuren burde nok venstre delen omfatte flere år, se figur 3. I tallene i figur 15 er det et sesongmønster som kommer klart fram når det er såpass mange år i diagrammet. Det er ikke noe tilsvarende sesongmønster i tallene i figur 2, og det kan av den grunn tas med færre år.

Figur 14. KURVEDIAGRAM MED FORSTØRRELSE AV DE SISTE ÅRENE



K i l d e : Statistisk ukehefte.

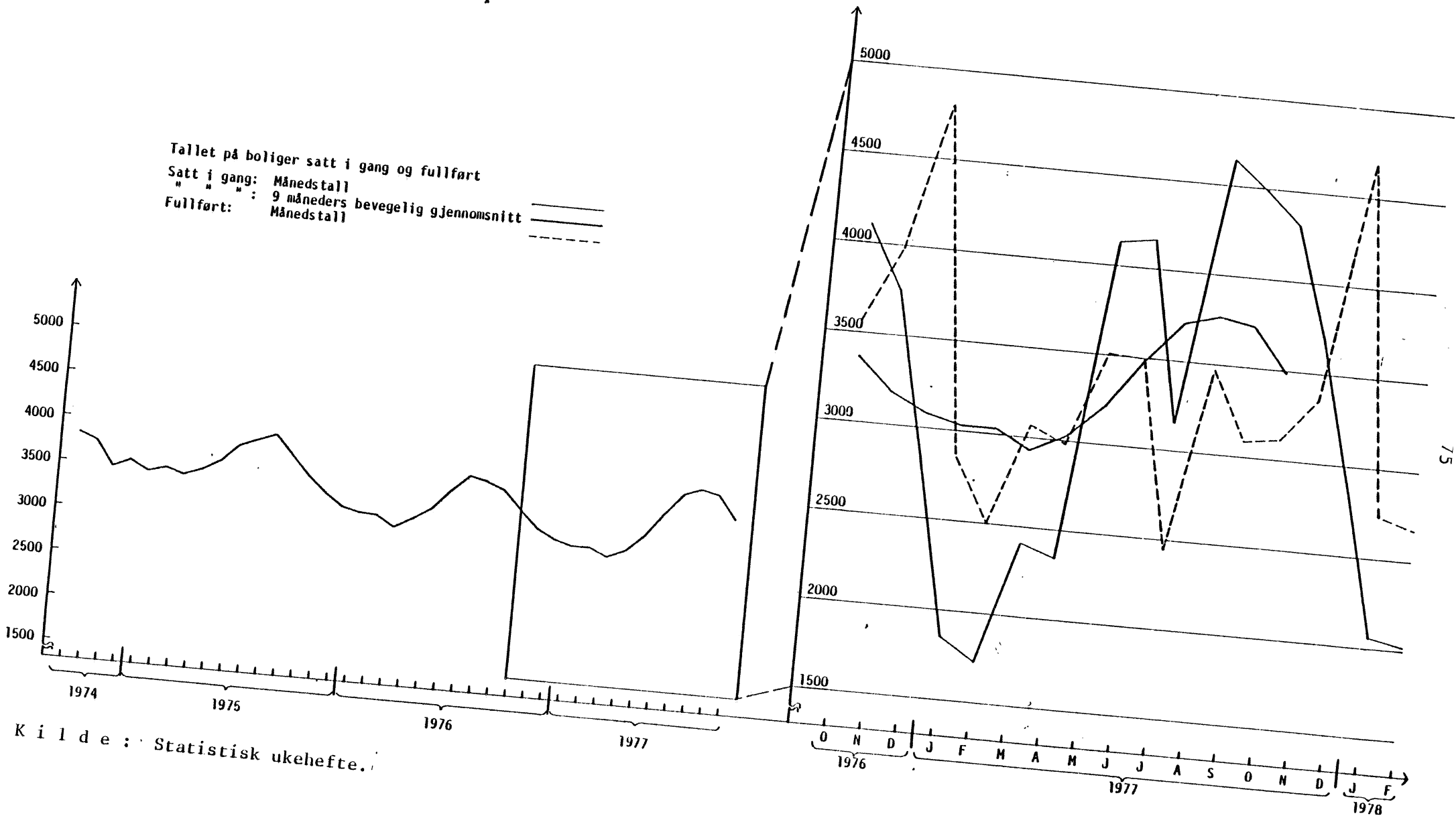
Figur 15. KURVEDIAGRAM MED FORSTØRRELSE AV DE SISTE ÅRENE.

Tallet på boliger satt i gang og fullført

Satt i gang: Månedstall

" " : 9 måneders bevegelig gjennomsnitt

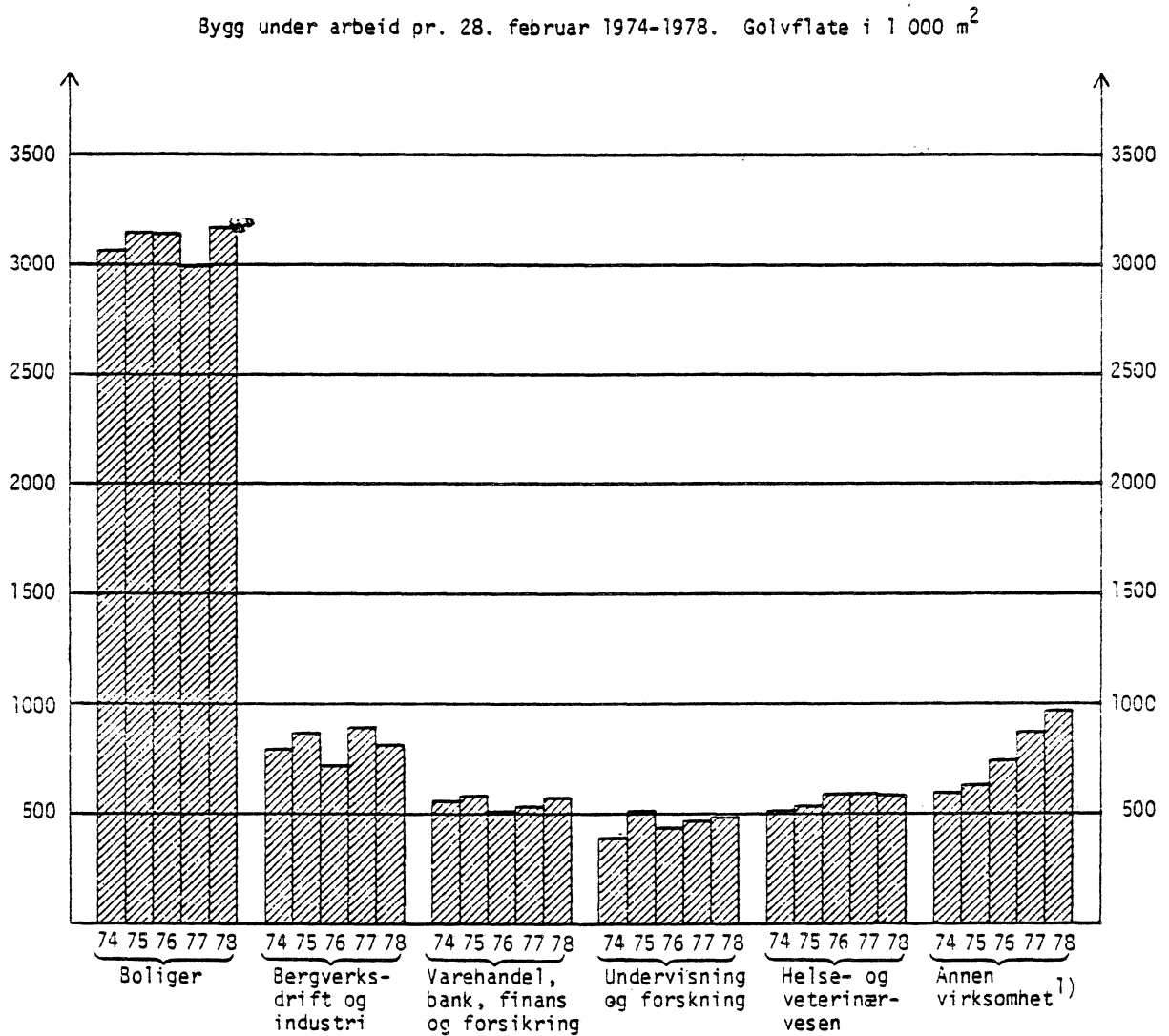
Fullført: Månedstall



Kilde: Statistisk ukehefte.

Stolpediagrammer er, som navnet sier, et diagram med stolper. Det finnes mange forskjellige slags stolpediagrammer, og i figur 16 har vi tegnet et enkelt stolpediagram med stående stolper. Av og til er stolpene liggende, av og til er de samlet i grupper som i figur 16, og av og til enkeltstående stolper. Liggende stolper, se figur 17, har den fordelen at vi kan plassere lengre horisintalt lesbare tekster tett ved hverandre. Samler vi stolpene i grupper, vil det kunne hjelpe på lesbarheten å skille mellom stolpene ved hjelp av skravering, fargelegging e.l.

Figur 16. STOLPEDIAGRAM



1) Omfatter ikke bygg for jordbruk, skogbruk og fiske.

Kilde: Statistisk ukehefte.

En spesiell form for stolpediagram er histogram¹⁾. Et histogram er et stolpediagram hvor stolpene gjengir en prosentfordeling (en frekvens-, hyppighets- eller tetthetsfordeling). Et slikt histogram er de velkjente befolkningspyramidene hvor befolkningen er fordelt etter alder og kjønn.

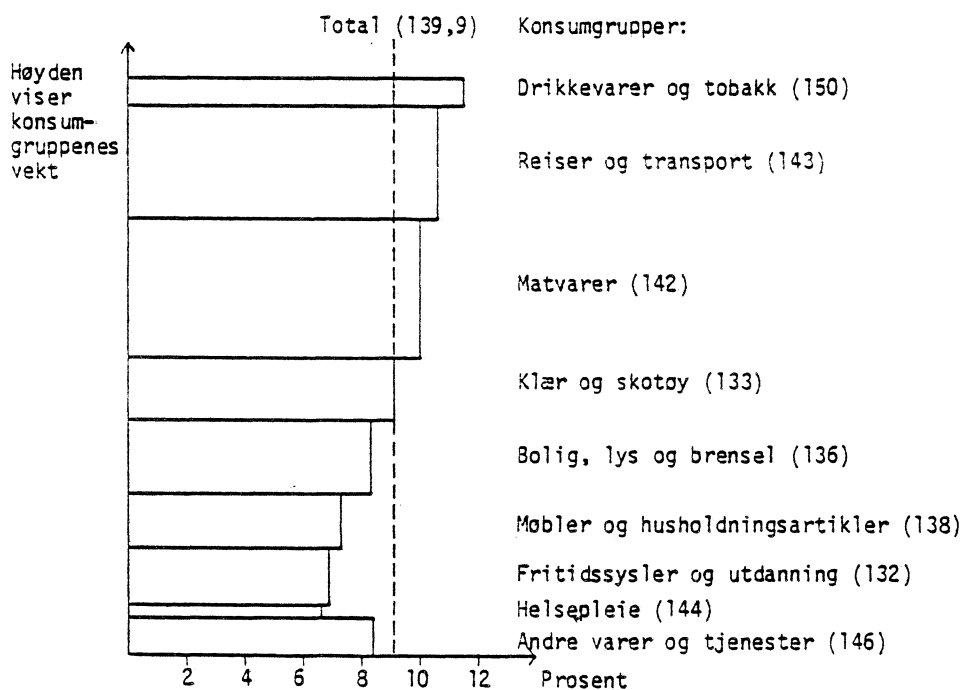
Når man skal velge mellom kurvediagram og stolpediagram for å illustrere en tabell, bør man huske at stolpediagram oftest tar lengre tid å tegne og at et kurvediagram forutsetter at den vannrette skalaen (x-aksen) er en kontinuerlig variabel.

Flatediagram illustrerer forskjeller i flater mellom klasser av verdier på ett eller flere kjennetegn. I et stolpediagram er det kun høyden på stolpene som sammenliknes, mens det i et flatediagram er flateinnholdet som sammenliknes. Dermed kan man i samme diagram anskueliggjøre større forskjeller enn i et stolpediagram, hvis vi da ikke velger en slags kombinasjon mellom stolpe- og flatediagram, se figur 17.

Figur 17. STOLPEDIAGRAM (FLATEDIAGRAM), LIGGENDE STOLPER

Konsumprisindeksen pr. 15. februar 1978 (1974 = 100)

Prosentvis endring i totalindeksen og de enkelte delindekser siste 12 måneder
(Indekstallene for februar 1978 i parentes. 1974 = 100)

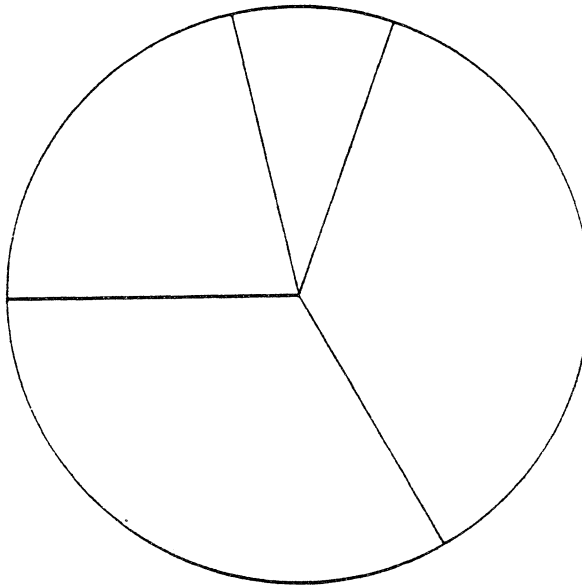


Kilde: Statistisk ukehefte.

1) Noen lærebokforfattere mener at et histogram er et flatediagram, dvs. et diagram hvor arealet er proporsjonalt med hyppigheten.

En spesiell form for flatediagram er det såkalte sektordiagram som viser prosentfordelinger. Et sektordiagram er altså en sirkel som deles opp i sirkelflater. Hele sirkelens flateinnhold er altså 100 prosent. Et slikt sektordiagram er meget oversiktlig, såframt verdiene på kjennetegnet vi vil illustrere deles opp i få klasser. De enkelte bitene bør skraveres, og prosentandelene kan ofte med fordel føres på figuren.

Figur 18



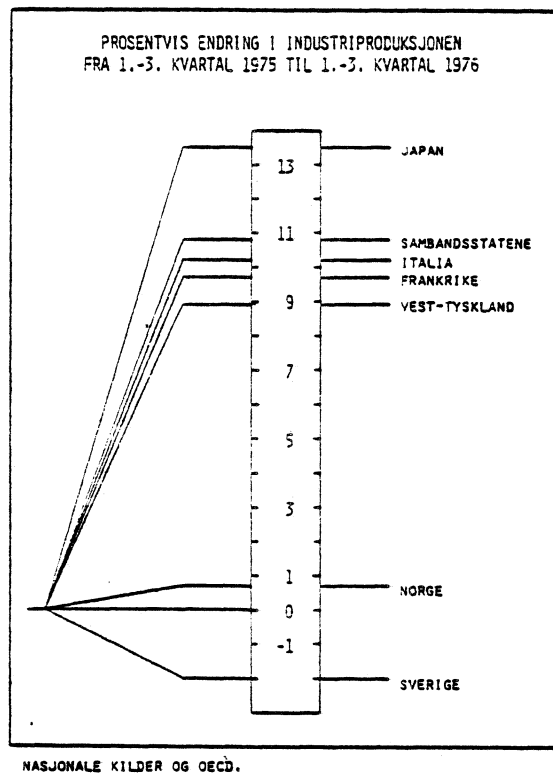
Volumdiagram bruker man når man vil sammenlikne klasser av kjennemerker med hverandre hvor størrelsesforskjellene er meget store. Svært mange mennesker har vanskeligheter med å lese slike tredimensjonalefigurer, og en bør derfor være varsom med å bruke dem.

Punktdiagram bruker man når man vil se på sammenhengen mellom to variable (to kjennemerker), f.eks. mellom inntekt og forbruk (inntekt langs x-aksen og forbruk langs y-aksen).

Trekantdiagram har nært slektskap med punktdiagrammet, men det viser variasjoner langs tre variable (de tre hjørnene i en trekant) i stedet for langs to, som i punktdiagrammet. Trekantdiagrammet er vanskelig å lese.

Fjærdiagram er en diagramtype som særlig er brukt i økonomisk utsyn, se figur 19.

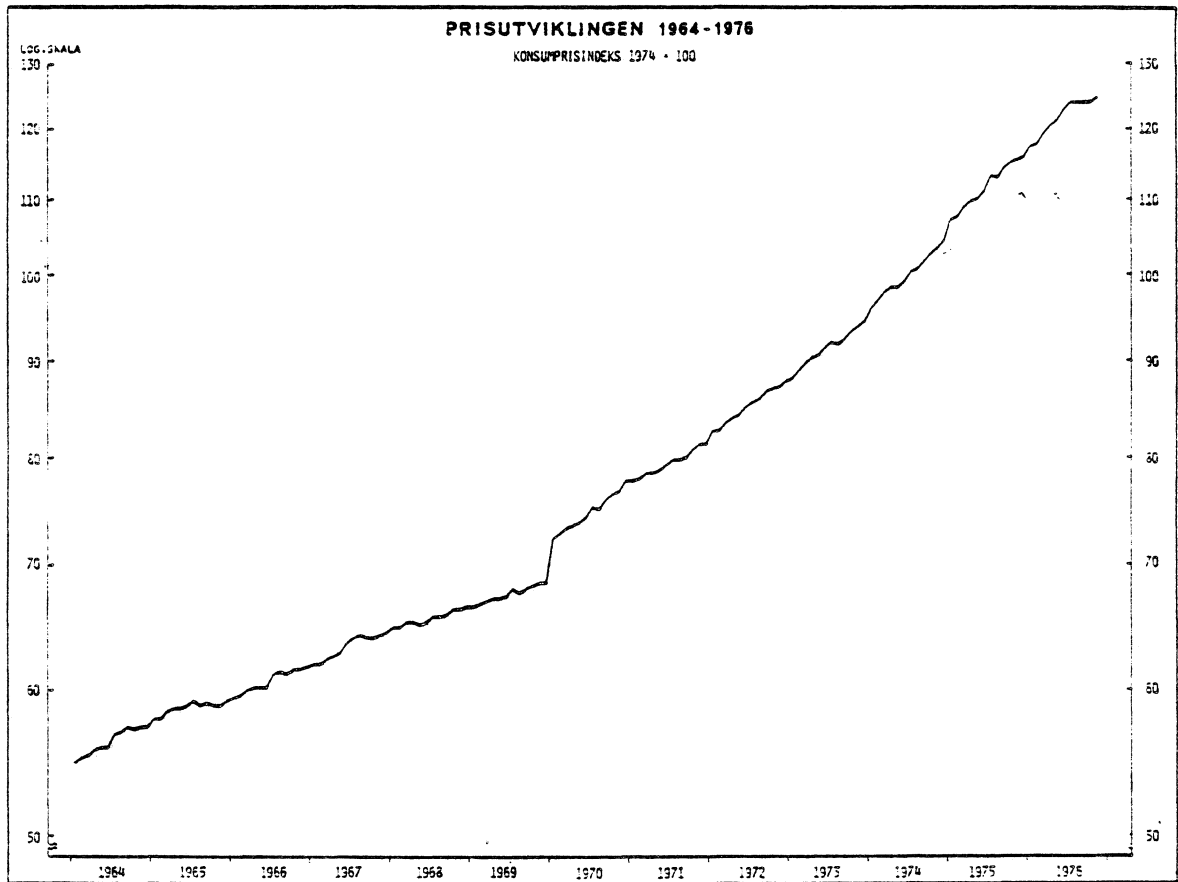
Figur 19. FJERDIAGRAM



Kilde: Økonomisk utsyn.

Diagramtypen brukes til å framstille prosentvise endringer i en periode, men den kan lett virke misvisende fordi avstanden fra det felles punktet til akse vil påvirke inntrykket av forskjeller. Tilsvarende innvendinger kan selvfølgelig også rettes mot andre diagramtyper: Valg av målestokk kan lett misvise leseren. Av og til brukes logaritmisk skala, se figur 20, og en slik skala vil være problematisk å lese og forstå for svært mange.

Figur 20. KURVEDIAGRAM MED LOGARITMISK SKALA

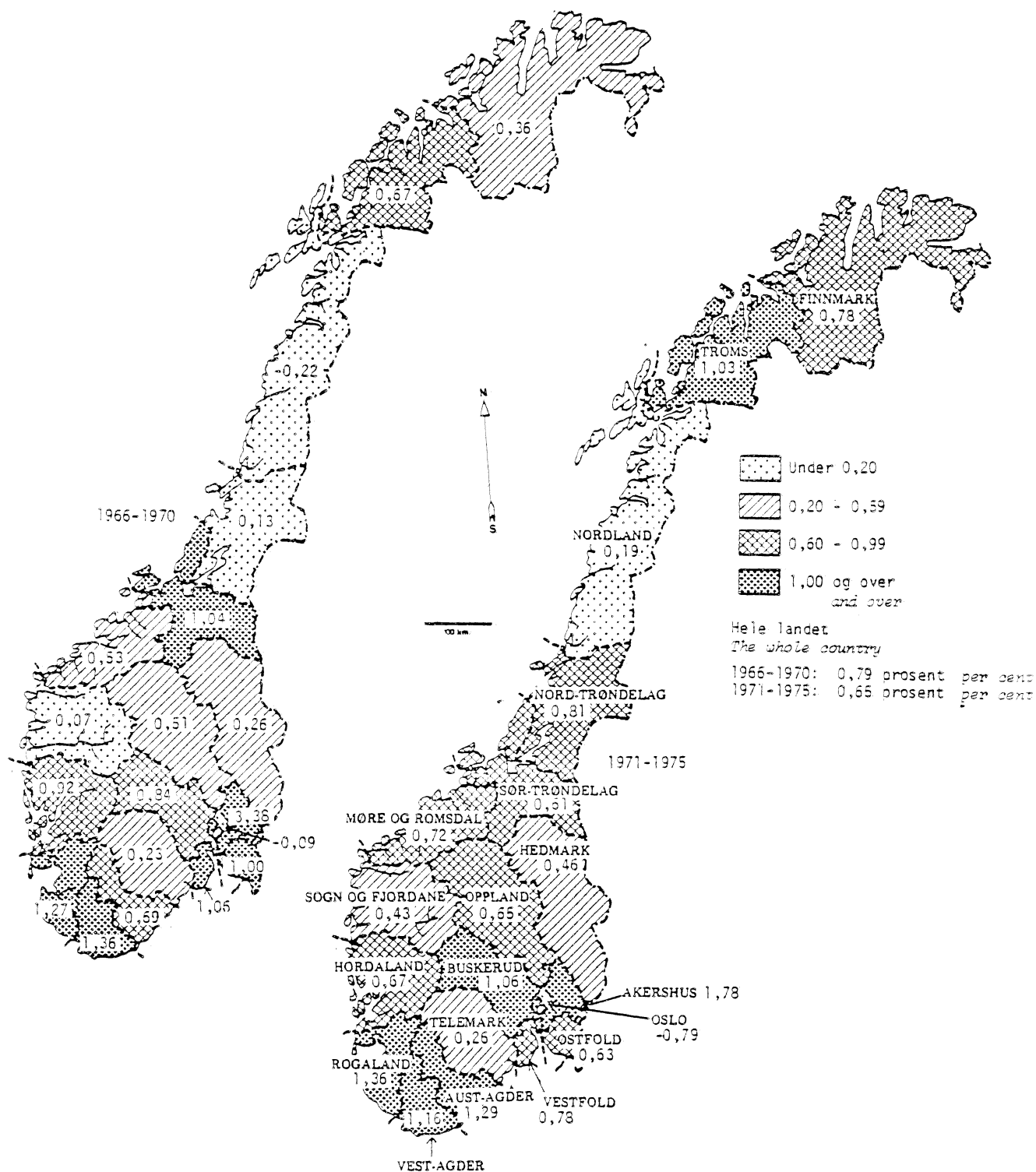


Kilde: Økonomisk utsyn.

Kartogram er at man på et kart, gjerne stilisert, tegner inn et fåtall geografiske forskjeller, f.eks. befolkningstetthet, biltetthet osv. For å illustrere forskjeller i tettheter, bruker en ofte forskjellige farger eller skraveringer. Kartogram er en meget anskuelig måte å få fram geografiske forskjeller på, men det kan være en omstendelig affære å tegne slike figurer som figur 21 og 22.

Figur 21. KARTOGRAM

Årlig folketilvekst i prosent. Fylke. 1966-1970 og 1971-1975 Annual percentage increase of population. County. 1966-1970 and 1971-1975

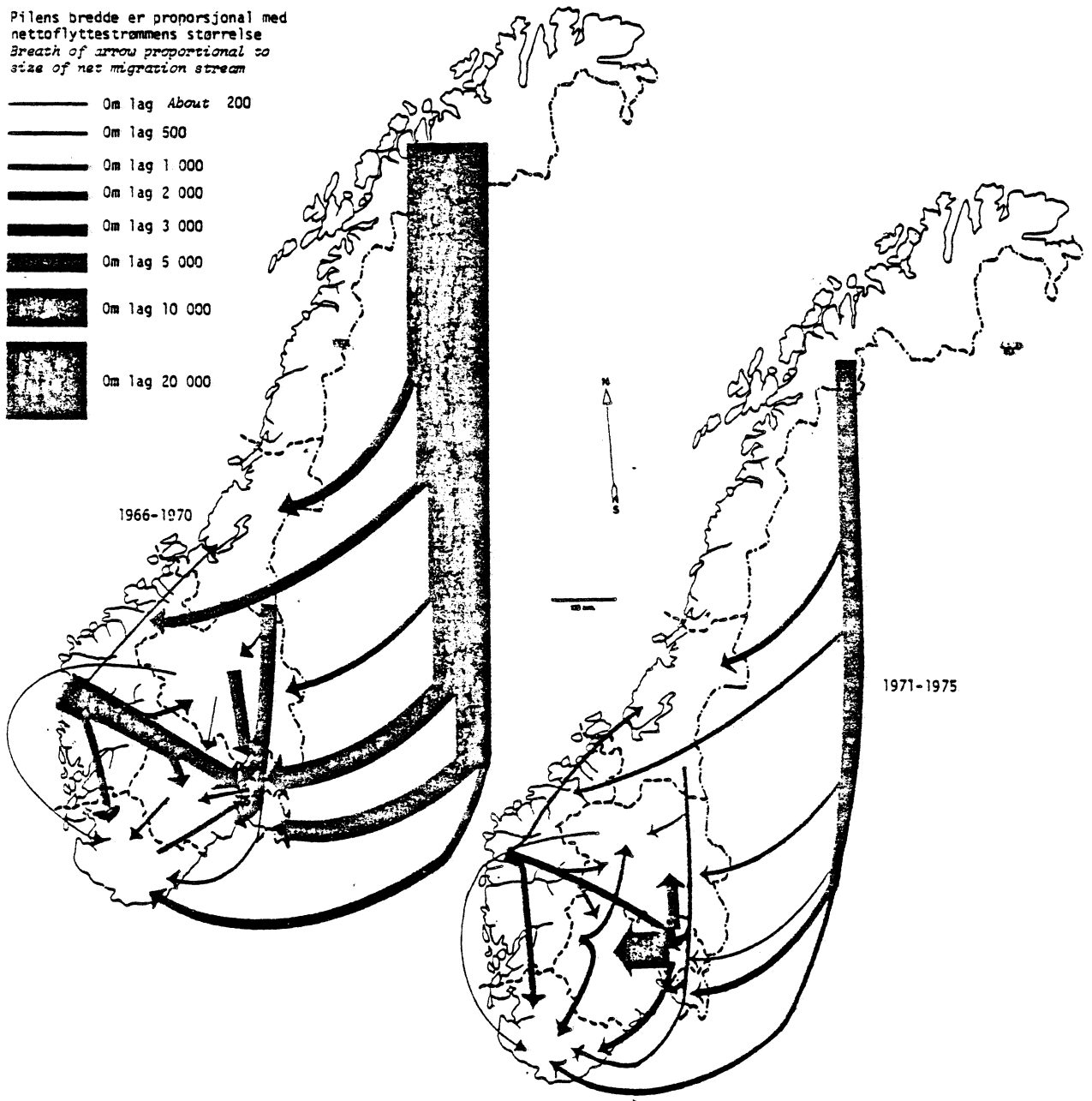
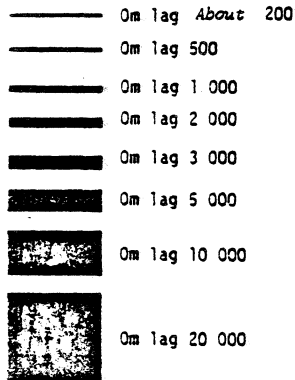


Kilde: Folkemengdens bevegelse. Oversikt 1971-1975.

Figur 22. KARTOGRAM

Nettoflyttestrømmene mellom landsdeier. 1966-1970 og 1971-1975 *Net migration streams between regions. 1966-1970 and 1971-1975*

Pilens bredde er proporsjonal med
nettoflyttestrømmens størrelse
*Breadth of arrow proportional to
size of net migration stream*



Kilde: Folkemengdens bevegelse. Oversikt 1971-1975.

B. Utregning av statistiske mål, gjennomsnitts- og spredningstall

Ved hjelp av gjennomsnitts- og spredningstall komprimerer vi en tabell ytterligere. Middeltall er gjerne fellesbetegnelsen på (det aritmetiske) gjennomsnitt, medianen, typetallet, det geometriske gjennomsnitt og det harmoniske gjennomsnitt. De to siste gjennomsnittstallene brukes sjelden, og en mener oftest med gjennomsnittet det aritmetiske gjennomsnitt.

I faget teoretisk statistikk behandles disse middeltallene mer inngående, og vi skal derfor ikke komme så mye inn på dem her.

Man bruker ofte slike middeltall når man ønsker å sammenlikne to masser av enheter med hverandre. En må da huske på at selv om sammenlikningsmålene er like (f.eks. gjennomsnitt), vil ikke det si at sammenlikninger kan foretas uten videre. Ved sammenlikninger av to datamaterialer må en vurdere om materialene, dvs. masse, enheter og kjennemerker, gir grunnlag for dette.

Spredningstall er tall som sier oss noe om spredningen av observasjonene (kjennemerkeverdiene). Ligger de tett samlet (om et eller annet middeltall), eller spredt langt utover? Spredningstall er slike tall som variasjonsbredden, fraktilene, (den empiriske) variansen og (det empiriske) standardavviket.

Endelig har vi visse tall som sier noe om sammenhengen mellom to (eller flere) variable, nemlig kovariansen og korrelasjonskoeffisienten. Korrelasjonskoeffisienten må ligge mellom verdiene pluss én og minus én. Vi snakker om negativ og positiv korrelasjon. Høy positiv korrelasjon har vi f.eks. når korrelasjonskoeffisienten er nær pluss én.

Slike empiriske mål som vi her har oppramset, bør ikke brukes hvis man ikke vet hva de står for. I faget teoretisk statistikk lærer man det. Boken Praktisk statistikk for samfunnsvidenskapene av Fisker, Nørgård, Thygesen, Wedebye og Østerlund (Akademisk Forlag, København 1977) gir en rekke enkle eksempler på bruk av gjennomsnitts- og spredningstall.

C. Indekser

I stedet for å si at et tall er 15 prosent større enn et annet tall, kan man angi forholdet mellom tallene som 115 til 100, altså at det ene tallet settes lik 100. Så snart man gjør det, pleier man å kalle tallene for indekstall.

Indekstall er vanlig å bruke ved sammenlikninger over tiden, tidsserier, men det brukes også ved andre relative sammenlikninger. Vanligvis mener man med en indeksserie en serie tall som ikke bare inneholder en omregning av en serie absolutte tall, men indeksserien er også et hjelpemiddel for å sammenfatte utviklingen for flere ulike hendelser som har ulike benevninger. Eksempelvis kan man ønske å sammenlikne arbeidsløshet (antall arbeidsløse) med lønnsutviklingen, trafikken på veiene (trafikkintensitet) med antall ulykker osv.

1. Konsumprisindeksen

Den viktigste og mest kjente av Statistisk Sentralbyrås indekser er konsumprisindeksen¹⁾. Selv om Byrået utarbeider svært mange andre indekser, både verdiindekser (indekser som viser endringer i en verdistørrelse over en tidsrekke), volumindekser og andre prisindekser, er bare konsumprisindeksen omtalt her.

De to hovedkomponentene som inngår i utregningen av konsumprisindeksen er mengdegrunnlaget som angir den betydning eller vekt som prisendringene for den enkelte vare tillegges i indeksen, og prisgrunnlaget som angir prisendringene for de enkelte varene.

Vektgrunnlaget (mengdegrunnlaget) har bygget på tall henholdsvis fra forbruksundersøkelsen i 1958, i 1967 og i 1973.

Prisgrunnlaget som skal samles inn hver måned, kan av naturlige grunner ikke omfatte prisene på alle merker, modeller o.l. for varer som inngår i forbruket i private husholdninger. F.eks. finnes det visstnok hele 70 forskjellige typer/merker husholdningspapir. Det er derfor valgt ut ca. 770 spesifiserte varer - representantvarene - som til sammen representerer prisendringene for hele spekteret av varer og tjenester som inngår i private husholdningers forbruk.

1) Beskrivelsen er hentet fra Statistisk Sentralbyrås Håndbøker nr. 32, utgitt i september 1977.

På grunn av endringer i forbruksvanene vil vektgrunnlaget etter noen tid ikke gi tilstrekkelig nøyaktig uttrykk for den aktuelle sammensetning av forbruket. Før avvikene blir så store at de er av betydning for indeksen, må vektgrunnlaget ajourføres. For å skaffe grunnlag for ajourføring av vektgrunnlaget er det gjennomført forbruksundersøkelser.

Utvalget av representantvarer kan ikke holdes uforandret i lang tid. Den relative betydning av varene endrer seg, og dette kan føre til at noen av representantvarene kommer til å spille mindre rolle i forbruket, mens nye varer som ikke er med i prisgrunnlaget, øker i betydning. Representantvareutvalget revideres nå løpende hvert halvår.

Vektgrunnlaget

Byrået er blitt stående ved å bruke tall fra forbruksundersøkelsene som mengdegrunnlag (vekter) i indeksen, nå resultater fra forbruksundersøkelsen 1973.

Forbruksundersøkelsen 1973 ble, i likhet med liknende tidligere undersøkelser, gjennomført som utvalgsundersøkelse etter den representative metode.

Parallelt med forarbeidene til forbruksundersøkelsen 1967 etablerte Byrået en permanent intervjuorganisasjon. Intervjuorganisasjonen har stått for innsamlingen av grunnmaterialet til forbruksundersøkelsene. I tillegg til de krav forbruksundersøkelser stiller til intervjuorganisasjonen er det bl.a. tatt hensyn til at den skulle være egnet til å foreta den løpende prisinnsamlingen for konsumprisindeksen.

Utvalget til forbruksundersøkelsene i 1967 og 1973 som omfattet ca. 4 700 private husholdninger, ble trukket slik at alle husholdninger i landet ble gitt samme sannsynlighet for å komme med i undersøkelsen.

Grunnmaterialet til forbruksundersøkelsen 1973 ble i likhet med 1967-undersøkelsen innhentet ved regnskapsførsel og intervju. Forskjellen er at i 1967 førte husholdningene regnskap i en måned, mot siste gang i en 14-dagersperiode. Etter at husholdningene var ferdig med regnskapsføringen ble de i 1973-undersøkelsen intervjuet bl.a. om kjøp av visse større varige forbruksvarer i løpet av den forutgående 12-månedersperioden, mens det til 1967-undersøkelsen ble innhentet opplysninger for kalenderåret.

Visse poster som faller inn under privat konsum er ikke tatt med i indeksgrunnlaget. Dette gjelder tjenester fra finansinstitusjoner, kontingenter til fagforeninger og andre foreninger, gaver o.l., skatter og trygder.

I tabell 1 er tatt inn det tidligere og det nåværende vektgrunnlaget for konsumprisindeksen.

Tabell 1. Oversikt over tosifrede konsumgrupper og vektorer i konsumprisindeksen

	Vektorer 0/00	
	1968=100	1974=100
Total	1 000,0	1 000,0
0 MATVARER	296,1	240,7
00 Mjøl, gryn og bakervarer	29,9	24,0
01 Kjøtt, kjøttvarer og flesk	75,8	62,7
02 Fisk og fiskevarer	23,8	19,1
03 Mjøl, fløte, ost og egg	56,6	41,5
04 Spisefett og spiseoljer	13,7	9,9
05 Grønnsaker, frukt og bær	44,2	37,4
06 Poteter og varer av poteter	11,0	8,5
07 Sukker	3,8	5,2
08 Kaffe, te, kakao og kokesjokolade	16,4	13,2
09 Andre matvarer	20,9	19,2
1 DRIKKEVARER OG TOBAKK	40,9	49,5
11 Drikkevarer	20,7	29,6
12 Tobakk	20,2	19,9
2 KLÆR OG SKOTØY	123,3	105,9
21 Bekleddningsartikler	86,9	77,6
22 Tøyer og garn mv.	14,7	10,6
23 Skotøy og skoreparasjoner	21,7	17,7
3 BOLIG, LYS OG BRENSEL	114,8	129,4
31 Bolig og vedlikeholdsutgifter 1)	71,4	94,0
32 Lys og brensel	43,4	35,4
4 MØBLER OG HUSHOLDNINGSARTIKLER	97,1	94,0
41 Møbler og golvtepper mv.	24,5	30,1
42 Tekstiler og utstyrsvare mv.	15,5	13,8
43 Komfyrer, kjøleskap og annet elektrisk utstyr	23,3	15,8
44 Kjøkkenredskap, glass, dekketøy mv.	10,1	10,2
45 Diverse husholdningsartikler og tjenester	18,7	18,2
46 Leid hjelp til hjemmet	5,0	5,9
5 HELSEPLEIE	19,2	23,2
51 Helsepleie	19,2	23,2
6 REISER OG TRANSPORT	161,9	195,9
61 Kjøp av egne transportmidler	63,3	63,4
62 Drift og vedlikehold av egne transportmidler 1)	57,5	82,8
63 Bruk av offentlige transportmidler	28,5	33,5
64 Porto, telefon og telegrammer	12,6	16,2
7 FRITIDSSYSLER OG UTDANNING	85,2	97,8
71 Utstyr og tilbehør, inkl. reparasjon	37,6	44,1
72 Offentlige forestillinger og andre tjenester	20,8	28,3
73 Bøker, aviser og tidsskrifter	20,4	21,8
74 Skolegang	6,4	3,6
8 ANDRE VARER OG TJENESTER	61,5	63,6
81 Personlig hygiene	18,3	21,0
82 Andre varer, ikke nevnt foran	12,0	12,8
83 Utgifter på restauranter, hoteller, selskapsreiser mv.	31,2	29,8

1) I forbruksundersøkelsen 1973 er eiendomsskatt og motorvognskatt tatt med under skatter og trygder.

Prismaterialet

Prisgrunnlaget for indeksberegningene er prisene på den enkelte representantvare. Utvalget av representantvarer revideres kontinuerlig. Hvert annet år oversender Byrået bransjeorganisasjoner o.l. utvalget av representantvarer til uttalelse og ber om eventuelle forslag til revisjon. Siden 1974 er valgt som basisår, ble en del av dette arbeidet utført i 1973. En større ajourføring av representantvareutvalget enn vanlig ble imidlertid utført omkring årsskiftet 1974-1975.

For boligkomponenten, bortsett fra vedlikehold, ble prisgrunnlaget for indeks med 1959=100 samlet inn ved en spesiell årlig boligundersøkelse. Et utvalg på vel 8 000 husholdninger (svarprosent ca. 70) ble spurt om utgift til husleie, renter av lån til eget hus eller leid leilighet, forsikring, skatter og avgifter til eget hus. Det vesentligste problem i denne forbindelse var knyttet til data for faktisk utgift til eget hus. Etter en inngående vurdering av forholdene, kom Byrået (i 1969) til at en i prisgrunnlaget for indeks med 1968 = 100 (som for indeks med basis før 1959 = 100) burde gå over til å nytte endring i husleie som indikator (prisgrunnlag) også for endring for eget hus. Byrået har seinere ikke funnet grunnlag for å endre praksis for innhenting av det kvartalsvise prismaterialet for boligkomponenten i konsumprisindeksen.

Prismaterialet for visse varer blir innhentet av Byrået fra visse sentrale organisasjoner (landbruket) og enkelte bedrifter. Dette gjelder først og fremst varer med ens prisutvikling for hele landet. For de andre varene blir prisene til det ajourførte beregningsgrunnlaget samlet inn hver måned gjennom Byråets intervjuorganisasjon. Fra og med januar 1975 tok Byråets intervjuorganisasjon i bruk en ny utvalgsplan. I første trinn i utvalgsplanen er landet delt i 5 deler:

1. Oslo - Akershus
2. Resten av Østlandet
3. Sørlandet - Vestlandet
4. Møre - Trøndelag
5. Nord-Norge

I tillegg er Oslo, Bergen og Trondheim tatt ut som egne områder, i alt 8 såkalte markeder. Utvalget er konsentrert til om lag 100 områder (ca. 120 kommuner), som er valgt ut etter utvalgsplanen for Byråets intervjuorganisasjon og hvor Byrået har (ca. 300) intervjuere. Bedriftsutvalget, dvs. antall butikker som er med, er ca. 1 400.

Indekstall for den enkelte representantvare for landet beregnes ved at relativ prisendring i hvert av de 8 markedene veies med omsetningen i detaljhandelen i de respektive markedene. Omsetningstallene er fra varehandelsstatistikk 1973.

For sterkt sesongpregede matvarer brukes de løpende priser for disse varene i indeksberegningene. Utenom sesong brukes gjennomsnittsprisen for foregående sesong.

Prisene innhentes pr. den 15. i måneden. Butikkene skal notere den pris de tar for varen denne dagen. For varer som er på billigsalg, dagens tilbud, sesongsalg o.l. når prisene noteres, skal de reduserte priser oppgis. Kvantumsrabatt blir derimot ikke fanget opp.

Indeksformel og basis

Konsumprisindeksen bygger på Laspeyres indeksformel. Indeksen skal følgelig som før gi uttrykk for forholdet mellom verdien regnet i månedens priser for de mengder som inngår i vektgrunnlaget og den tilsvarende verdi etter basisårets priser.

Indeksformelen kan også skrives:

$$I_t = 100 \sum a \cdot \frac{p_t}{p_0}$$

hvor a -ene er budsjettprosentene (de relative andelene av utgiftene til de enkelte varene) i vektgrunnlaget og p_t/p_0 er forholdet mellom pris på beregningstidspunktet (t) og prisen i basisåret (0) for de enkelte varer. Summeringen er foretatt over alle varer. Som nevnt er budsjettprosentene (vektene) beregnet på grunnlag av forbruksundersøkelsen 1973, og de relative pristall er beregnet på grunnlag av prismaterialet som samles inn hver måned for representantvarene og representantvarenes priser i basis.

I konsumprisindeksen er gjennomsnittsprisene for 1974 satt som basis for prissammenlikningene, dvs. gjennomsnittet for 1974 er satt lik 100.

Framtidig kontroll av vekt- og prisgrunnlaget

Forhold som fra tid til annen gjør det nødvendig å revidere beregningsgrunnlaget for konsumprisindeksen, er som nevnt endringer i forbruket som kan tilsi endringer av vektene i indeksen og endringer av forbruksvarene som kan være av betydning for utvalget av representantvarer.

I og med at grupperingen i konsumprisindeksen har kommet på linje med den grupperingen av privat konsum som vil bli lagt til grunn i nasjonalregnskapet, og at en fra og med 1973 også har årlige forbruksundersøkelser, har Byrået fått et betydelig bedre grunnlag enn tidligere for løpende kontroll av vektene i konsumprisindeksen.

Når det gjelder representantvarene, gir innsamlings- og bearbeidingsrutinene etter det nåværende opplegg gode holdepunkter for løpende vurdering av utvalget av representantvarer; for eksempel ved store endringer i antall prisnoteringer og/eller endringer i spredningen i prisoppgavene. Systemet tillater også løpende revisjon av utvalget, dvs. hvis en representantvare pga. kvalitetsendringer, nye modeller eller merker av varen har mistet noe av sin betydning, kan den skiftes ut og erstattes med en ny uten at dette forårsaker brudd i indeksen. Likeledes vil endringer som bringer prisutviklingen for en representantvare i utakt med prisbevegelsen for den konsumgruppe som den skal representere, kunne motvirkes på liknende måte. Opplegget for indeksen gir derfor i stor utstrekning grunnlag for beregning av en indeks som gir mest mulig korrekt bilde av prisutviklingen.

Levekostnader

Det problem en står overfor i konsumprisindeksberegninger kan formuleres slik: Hvis prisene endrer seg fra ett tidspunkt til et annet, hvilket utgiftsbeløp (eller hvilken inntekt) er da nødvendig for å opprettholde samme levenivå ("levestandard") som på det første tidspunkt?

Noen anerkjent metode for å måle levenivået finnes ikke, bl.a. fordi en rekke av de faktorer som de fleste vil regne for å være av betydning for levenivået, vanskelig lar seg måle statistisk. Det lar seg derfor ikke gjøre å måle nøyaktig kostnadene ved opprettholdelsen av et bestemt levenivå, og de praktiske konsumprisindeksberegninger må av den grunn få en mer begrenset målsetting.

Ved konsumprisindeksberegninger er problemstillingen derfor å beregne kostnadene for et bestemt forbruk av varer og tjenester på ett tidspunkt i forhold til kostnadene for det samme forbruk på et annet tidspunkt.

Nettoprisindeks

Det byr på betydelige statistiske (og teoretiske) problemer å beregne en nettoprisindeks, en prisindeks korrigert for avgifter og tilskott. Byrået har ikke foretatt beregninger av konsumprisindeksen korrigert for avgifter og tilskott. Først må nemlig prismaterialet, både for basisperioden (1974) og beregningstidspunktet renses for tilskott og avgifter. Dette er særlig vanskelig for de varer hvor avgiftene er lagt på varens råstoffer eller halvfabrikata. I alminnelighet kan en heller ikke regne med at alle indirekte skatter og subsidier fullt ut blir overveltet i prisene. Dessuten burde en prinsipielt også skaffe et forbruksgrunnlag som er tilpasset de priser en får når en tar bort avgifter og subsidier. Det vil ikke være nok bare å beregne hvor meget av det registrerte utgiftsbeløp som kan sies å være avgifter og tilskott. En må gå ut fra at dersom forbrukeren stod overfor en så betydelig endret prissituasjon, ville det være "lønnsomt" for ham å foreta en omlegging av forbruket (vektgrunnlaget).

Indeksklausuler

Mange kontrakter (avtaler) om økonomiske forhold har indeksklausuler. I den forbindelse er det ofte behov for å regne om poeng i konsumprisindeksen til prosent, og dessuten ved endring av sammenlikningsbasis for indeksen å regne om indekstall med en basis til indekstall med en annen basis. På neste side følger to eksempler på slike omregninger.

Eksempel på beregning av poengvis endring i konsumprisindeksen til prosentvis endring:

15. mars 1969 - 15. desember 1974

66,8 (indeks) 104,2 (indeks)

Stigning i poeng: 37,4

Stigning i prosent: $\frac{104,2 - 66,8}{66,8} \cdot 100 = \underline{56,0}$

Konsumprisindeksen har vært beregnet etter følgende basisår:

Til og med juli 1940: juli 1914 = 100 (Leveomkostningsindeks)

Fra januar 1939 til september 1950: 1938 = 100 (")

Fra januar 1950 til april 1962: 1949 = 100 (Levekostnadsindeks)

Fra januar 1960 til februar 1969: 1959 = 100 (Konsumprisindeks)

Fra januar 1968 til mars 1975: 1968 = 100 (")

Fra januar 1974: 1974 = 100 (")

Eksempel på omregning av indekstall med 1974 = 100 til indekstall med 1968 = 100:

15. april 1975 - Gj.snitt 1968

109,4 (indeks) 65,5 (indeks)

Indekstall 15. april 1975, 1968 = 100: $\frac{109,4}{65,5} \cdot 100 = 167,0$

7. PUBLISERING OG LAGRING AV DATA

A. Publisering

Det siste leddet i en statistisk undersøkelse er en analyse av resultatene. Denne analysen, som oftest ganske elementær, gjøres kjent for brukerne i en eller annen publikasjon. Analysen må være tilpasset brukerne, dvs. de må forstå den. Både valg av analysemetoder, dvs. hvordan materialet skal analyseres og hvilke resultater av analysen som skal publiseres, inngår i det vi vil kalle analysen. Når en publikasjon bare inneholder tabeller uten kommentarer, er det også resultatene av en analyse som offentliggjøres: Materialet er analysert ved hjelp av tabeller, og analytikeren har valgt ut de tabeller som best illustrerer eller belyser de problemstillinger som var bakgrunnen for innsamlingen av materialet. Ofte vil tabellene i en publikasjon også være ledsaget av en tekstlig analyse.

1. Publiseringsmåter og kommentarer

Byrået utgir om lag 180 ulike publikasjoner i året, av disse kommer nærmere 100 en gang pr. år, eller med regelmessige intervaller. Byrået gir ut 5 publikasjoner som kommer oftere enn en gang i året. Av periodiske publikasjoner har vi:

1. S t a t i s t i s k u k e h e f t e (SU), som utkommer hver uke. Ved årets slutt gis det ut et ekstra ukehefte som inneholder et alfabetisk register over de tabeller ukeheftet har inneholdt i løpet av året. Ukeheftet består av tabeller pluss en kort tekstlig kommentar til tabellene. Det trykkes et visst antall særtrykk av de forskjellige statistikkene som gjengis i ukeheftet.
2. N y e d i s t r i k t s t a l l (ND). Et hefte for hvert fylke kommer en gang i måneden. Heftene sendes ut i puljer på 4-5 fylker. Heftene er ment bl.a. å skulle oppdatere tabellene i Statistiske fylkeshefter (se seinere).
3. S t a t i s t i s k m å n e d s h e f t e (SM), som utkommer hver måned, inneholder nå bare tabeller. Tidligere når en statistikk ble omlagt, ble denne omleggingen beskrevet i et månedshefte; nå blir ukeheftet brukt til dette. Månedsheftet inneholder tabeller for de fleste områder der det foreligger måneds- eller kvartalsstatistikk, og det er også tatt med tilbakegående tall, tidsrekker.

4. M å n e d s h e f t e over u t e n r i k s h a n d e l e n (MU) gir månedlige oversikter over utenrikshandelen i form av tabeller.
5. K o n j u n k t u r t e n d e n s e n e utgis en gang i måneden og gir en verbal beskrivelse av konjunkturtendensene i utlandet og i Norge. Enkelte tabeller og diagrammer er vedlagt. Tabellene oppdateres gjennom ukeheftet.
6. N o r g e s o f f i s i e l l e s t a t i s t i k k (NOS) utgjør hovedtyngden av Byråets publikasjoner. I denne serien publiseres resultatene fra de løpende undersøkelsene og resultater fra undersøkelser som gjentas med visse mellomrom (f.eks. folketellingen). I publikasjonen Statistisk årbok samles hovedresultatene fra de aller fleste statistikkene som utarbeides i Byrådet, dessuten er det også med visse internasjonale oversikter. I Årboka gis det en del tilbakegående tall, men det er i publikasjonen Historisk statistikk (siste fra 1978, kommer ut hvert 10. år) at vi finner tall så langt tilbake som de overhodet finnes. Historisk statistikk inneholder dessuten korte tekstavsnitt med en oversikt over statistikutviklingen innenfor de enkelte fagområder. Statistiske fylkeshefter utgis også i serien NOS med 3-4 års mellomrom. Siste serie fylkeshefter kom ut i 1980.
7. A k t u e l l s t a t i s t i k k (AS) kan neppe kalles en publikasjon, fordi det som oftest bare er en side med hovedresultater fra en undersøkelse. Alle undersøkelser omtales først i Aktuell statistikk, og når undersøkelsen er omtalt der, er tallene frigitt, dvs. da kan resten av resultatene også opplyses til eventuelle brukere som ikke kan vente til publikasjonen(e) er ferdigtrykt. Aktuell statistikk går for det meste til våre nyhetsmedia, og de bruker ofte ordrett det som Byrådet skriver. Det blir sendt ut flere pressemeldinger nesten hver dag, og i løpet av et år sendes det ut mellom 450 og 500 pressemeldinger. I tillegg til de periodiske publikasjoner kommer seriene RAPP, ART, SØS og SA. Publikasjoner i disse serier kommer uregelmessig.
8. R a p p o r t e r (RAPP) er en ny publikasjonsserie, som omfatter analyser som antas å ha en begrenset leserkrets.
9. A r t i k l e r (ART) inneholder forholdsvis korte tekstlige analyser av statistiske materialer eller drøftelser av visse statistiske metoder eller andre mer teoretiske problemstillinger. Forfattere kan hentes fra kretsen utenfor Byråets ansatte.
10. S a m f u n n s ø k o n o m i s k e s t u d i e r (SØS) er større og oftest vitenskapelige avhandlinger. Forfatterne av disse studiene er ikke alltid Byråets egne ansatte.
11. S t a t i s t i s k e a n a l y s e r (SA) er en forholdsvis ny publikasjonsserie hvor resultatene fra statistiske undersøkelser presenteres i tekst og tabeller. Dessuten er det ofte slik at det statistiske materialet som omtales i de statistiske analysene, har gjennomgått en grundigere analyse enn det har gjort når det blir publisert i andre tabellpublikasjoner.
12. S t a n d a r d e r f o r n o r s k s t a t i s t i k k (SNS) vil heretter bli gitt ut som egen serie. Tidligere kom de i serien Statistisk Sentralbyrås Håndbøker (se 13). Standardene gir regler klassifisering av enheter i offisiell norsk statistikk.

Av helt spesielle publikasjoner som Byrået utgir, kan nevnes:

13. S t a t i s t i s k S e n t r a l b y r å s H å n d b ø k e r (SSH), dvs. oppslagsbøker egentlig beregnet for de ansatte som hjelpemidler i arbeidet.
14. S t a t i s t i s k e k o m m u n e h e f t e r ble utgitt i forbindelse med Folke- og boligtellingsen 1970 (og i 1960) for hver enkelt kommune. Tilsvarende hefter vil komme etter Folke- og boligtellingsen 1980, men vil ikke bli gitt ut som egen serie.

Kommentarene som gis til et statistisk materiale bør være tilpasset brukerne. Når man skal kommentere en tabell skriftlig, bør en bruke de reglene en har lært for å lese en tabell, fordi en skriftlig kommentar til en tabell henvender seg nettopp til de brukere som ikke alltid er vant til å lese tabeller. Vi bør altså ved vår kommentar "lese" tabellen for dem.

I v. Hofstens bok Praktisk statistikk er gitt omtrent følgende standard disposisjon for redigjøreelse av en statistisk undersøkelse:

1. Omslag og tittelsider
2. Innholdsfortegnelse, ofte med fortegnelse over tabeller og diagrammer også
3. Forord
4. Innledning som omfatter bakgrunnen for undersøkelsen, hvem som har finansiert den, osv.; dessuten vil det ofte være en fordel å omtale resultatene fra tidligere liknende undersøkelser
5. Formålet med undersøkelsen og presisering av omfang, enheter og definisjoner
6. Organiseringen av undersøkelsen med opplysninger om hvordan materialet er bearbeidd
7. Presentasjon og analyse av resultatene
8. Eventuelle bilag hvor bl.a. spørreskjemaet, komplisert matematikk, litteraturliste etc. er gjengitt

Avsnittene 1-4 er svært like fra undersøkelse til undersøkelse og behøver ingen kommentarer, mens avsnittene 5-8 ofte varierer fra en undersøkelse til en annen. FN's underkomité for statistisk sampling har kommet med følgende regler for punktene 5-8:

- a) Formålet med undersøkelsen må være helt klart
- b) En beskrivelse av det statistiske materialet. De geografiske områdene skal angis nøyaktig, likeså andre avgrensinger av den masse som undersøkelsen omfatter
- c) Detaljerte opplysninger om det innsamlende materialet. De(t) anvendte skjema bør gjengis, helt eller delvis; dessuten utfyllingsregler som er av særskilt betydning
- d) Innsamlingsmetoden (post, intervju etc.); dessuten bør hvis mulig også opplysninger om revisjon og bearbeiding av materialet gis
- e) Utvalgsmetoden må angis hvis undersøkelsen er en utvalgsundersøkelse. En bør da gi utvalgets størrelse, dekningsgrad, frafallsprosent etc.

f) Nøyaktigheten i de oppnådde resultatene. En bør søke å gi en redegjørelse for ulike feilkilder og gi et anslag på feilenes størrelse (numerisk)

g) Kostnadene bør helst gis både totalt og fordelt på de ulike deler av undersøkelsen

Det vil avhenge av hvem publikasjonen henvender seg til om en del av opplysningene under a)-g) bør plasseres i vedlegg eller i selve teksten. Detaljerte regler for innhold mv. i Byråets publikasjoner, som stort sett er i overensstemmelse med reglene foran, er gitt i Håndbok nr. 40¹⁾.

2. Trykkeriteknikk

Når man skal lage en publikasjon eller en annen redegjørelse for en statistisk undersøkelse, må en ha klart for seg hva som er mulig å gjennomføre (dvs. hvilke tekniske hjelpemidler som er til disposisjon) og hvor mye det koster å bruke de enkelte hjelpemidlene. En må vite kostnadene når en skal velge trykkmetoder.

Ofte vil man kunne velge mellom følgende fem metoder for mangfoldiggjøring:

1. Boktrykk
2. Offsettrykk
3. Stensilering
4. Kopiering
5. Maskinskriving med gjennomslag

Boktrykk

Teksten som skal trykkes blir vanligvis satt i bly. Satsen kan bygges opp av enkelttyper, hele linjer, eller den kan være støpt i ett.

Avtrykkene framstilles i boktrykkmaskinen ved direkte kontakt mellom blysetsen og trykkipapiret. Metoden gir skarpe avtrykk, men forarbeidet er tidkrevende og trykkhastigheten relativt lav.

Offsettrykk

Utgangspunkt her er en original som kan være framstilt ved hjelp av fotosats, skrivemaskin, printer (EDB), eller som avtrykk, f.eks. av blysets.

Tekstbildet på originalen overføres til en offsetplate, enten direkte eller via fotografisk film.

I offsetmaskinen tar offsetplaten opp sverte på de steder der det er tekst. Tekstbildet overføres fra platen til en gummiduk og derfra til trykkipapiret.

Ved bruk av gode originaler, f.eks. utskrift fra fotosetter, kan metoden måle seg med boktrykk i kvalitet.

Offsetmetoden har stadig vunnet terreng, bl.a. på bekostning av boktrykk.

Stensiltrykk

Som trykkmedium nyttes det en stensil. Stensilen gjøres reproduserbar ved å skrive på den med hånd eller maskin. Det kan også overføres tekst fra en original ved hjelp av en elektronisk metode. I alle tilfeller blir stensilens overflate brutt, slik at sverte kan trenge gjennom der det er tekst.

I stensilmaskinen produseres avtrykkene ved direkte kontakt mellom stensil og trykkipapir.

Metoden gir uskarpe avtrykk og blir stadig mindre brukt.

1) Statistisk Sentralbyrås Håndbøker 40: Regler for redigering og utstyr i publikasjoner fra Statistisk Sentralbyrå: Oslo 1978 (oppdateres løpende).

Kopiering

Ved kopiering skjer det for hver kopi en overføring fra original til papir, direkte eller indirekte.

Mest vanlige metoder i dag er en direkte metode basert på lysømfintlig papir (papir belagt med sinkoksyd) og den indirekte metoden, som er basert på det xerografiske prinsipp og som gir kopier på vanlig papir.

De mest avanserte kopiautomater arbeider etter sistnevnte metode, og kan produsere over 7 000 kopier i timen. For små opplag, ca. 15-500, utgjør de et alternativ til offsettrykk. For kopiering under ca. 15 pr. original nyttes mindre maskiner.

Gjennomslag

Bruk av karbonpapir muliggjør overføring av håndskrift og maskinskrift. I stedet for å legge inn karbonpapir kan baksiden av papiret være karbonisert. Det kan også nyttes et spesielt bestrøket papir, selvkopierende papir.

Kvaliteten av gjennomslagene avtar med antall eksemplarer og papirets tykkelse.

Kostnader

Valg av trykkmetode har betydning for kostnadene. Vel så viktig for totalkostnadene, er det at den enkelte trykksak blir produsert på maskiner som er egnet for det aktuelle sidetall, opplag og utstyr. Det er kostbart å trykke små opplag på store maskiner og store opplag på små maskiner. Enkelte trykksaker krever spesialmaskiner, f.eks. trykksaker som skal viderebehandles i en printer. Spesielle kvalitetskrav vil gjerne medføre større kostnader. Særlig for små opplag vil det ofte være spørsmål om å nøye seg med redusert, akseptabel kvalitet.

Etter selve trykkingen følger gjerne et ferdiggjøringsarbeid i form av falsing, stifting, binding osv. Hvordan ferdiggjøringen skal utføres, har betydning for kostnadene, og vil også kunne påvirke valg av utstyr for trykkingen.

Av andre faktorer som har betydning for kostnadene er manuskriptets kvalitet og produksjonstidens lengde.

Det er altså en rekke faktorer som har betydning for hva en trykksak vil koste. Produksjonsmåten bør overveies på et tidlig stadium under planleggingen av trykksaken, som regel før utarbeidingen av manuskriptet tar til.

3. Spesialservice

Hensikten med det arkivstatistiske system er at en (letterer) skal kunne yte tjenester til brukere hvor tjenesten (dvs. den statistiske informasjonen) kan være av spesiell karakter. Med tjeneste av spesiell karakter forstår vi informasjon som ikke er av generell interesse og derfor ikke publiseres. Det er utarbeidd en oversikt over de spesialtjenester Byrået yter og hva de koster (Spesialservice fra Statistisk Sentralbyrå). Disse tjenestene faller naturlig i fire grupper:

- Gratis opplysningstjeneste
- Abonnement på regelmessige oppgaver, i første rekke utenrikshandelsstatistikk, produksjonsstatistikk og regnskapsstatistikkens nøkkeltall
- Spesialoppdrag, utkjøring eller oppstilling av spesielle tabeller
- Fotostatkopier av enkeltsider fra Byråets publikasjoner og maskintabeller

Distribusjon av Byråets publikasjoner

Byråets kunder kan deles i to hovedgrupper etter distribusjonsmåten:

1. Offentlige etater (stat, fylke, kommune) bibliotek og undervisningsinstitusjoner

2. Andre

Publikasjoner til gruppe 1 sendes i posten. Informasjonsavdelingen har et Forsendelsesregister (F-reg) med noe under 20 000 adresser. F-reg er lagt opp på EDB, slik at adresseslipper kan kjøres ut for ulike mottakergrupper. Slippene klistres på konvoluttene i en adresseringsmaskin, og de fleste publikasjonene legges i konvoluttene ved hjelp av en maskin. Dette utføres av Gruppe for forsendelse (Trykningskontoret).

Informasjonsavdelingens Opplysningskontor tar imot retting og nybestilling til F-reg og betalt abonnement på Statistisk månedshefte, Statistisk ukehefte og Månedssstatistikk over utenrikshandelen. Dessuten tar kontoret imot bestilling på gratispublikasjonene Nye distriktstall, Konjunkturtendensene og Særtrykk fra Statistisk ukehefte. Opplysningskontoret har også publikasjoner for kontantsalg (ved fram møte).

Distribusjonen til gruppe 2 andre går gjennom Forlagsentralen.

Bokhandlerne bestiller publikasjonene via Forlagsentralen, som er distributør for Byråets kommisjonærer: Universitetsforlaget og Aschehoug & Co. A/S.

Publikasjoner i seriene NOS, SA, SØS, ART og RAPP skal det tas betaling for, dersom bestilleren ikke er en offentlig etat (se neste avsnitt). Byrådet kan dessverre ikke ta imot telefoniske eller skriftlige bestillinger på publikasjoner som det skal betales for. Unntak er fast abonnement på Statistisk ukehefte, Statistisk månedshefte og Månedssstatistikk over utenrikshandelen.

Universitetsforlaget, Avd. for offentlige publikasjoner, Tollbugt. 31, Oslo 1, tar imot bestillinger på Byråets publikasjoner - pr. telefon 41 24 35 - eller pr. brev. Publikum kan også vises til bokhandlerne.

Hvem får publikasjoner gratis?

Offentlige kontorer og institusjoner knyttet til stats-, fylkes- eller kommuneadministrasjon, offentlige bibliotek, aviser, norske ambassader i utlandet, universiteter og høyskoler får de publikasjoner de ønsker i et rimelig antall (men ikke klassesett til bruk i undervisning). Dessuten kan organisasjoner, politiske partier og skoler få fast gratisabonnement etter bestemte regler, som er fastlagt av Administrasjonsavdelingen og håndheves av Informasjonsavdelingen (Markedsføringsgruppa).

B. Lagring

I Byrådet opererer man med fire lagrings- eller arkivordninger:

H å n d a r k i v - lagring på den enkelte saksbehandlers "hånd", dvs. på hans/hennes rom

K o n t o r a r k i v - et sentralt arkiv for hele kontoret

S e n t r a l a r k i v - et bortsettingsarkiv i kjelleren

B å n d a r k i v e t - arkivet av magnetbånd (se neste kapittel om dokumentasjon av data)

Muligens kan det også være riktig å inkludere Biblioteket i denne arkiv- og lagringsoversikten.

Et arkiv er ikke et sted å bli "kvitt" et statistikkmateriale, og det er derfor utarbeidd kontroversive og sentrale instruksjoner for når et statistikkmateriale skal arkiveres og hvor lenge det skal arkiveres på de forskjellige stedene (se neste kapittel).

Det må også nevnes at Riksarkivet overtar visse sentrale statistikkdokumenter som man har vurdert har historisk interesse.

Arkiveringsteknikken er nå i ferd med å revolusjoneres, idet mikrofotografering vil gjøre det mulig å lagre nærmest ubegrensede mengder. Begrensningene vil heretter mer ligge i fotokostnadene og kostnadene ved opprettelse, og spesielt ajourføring av lageroversikter (kartotek e.l.).

C. Dokumentasjon av data¹⁾

Med dokumentasjon av data er det i denne sammenheng først og fremst tenkt på beskrivelser eller definisjoner av individualdata og aggregerte data beregnet på statistikkbrukerne (både i og utenfor Byrået) og statistikkprodusentene (fagkontorene i Byrået). Det er i mindre grad tenkt på den dokumentasjon som brukes av dem som utfører databearbeidingen (Systemkontoret og Driftskontoret). Dokumentasjon av data er en del av den komplette dokumentasjon av en statistikk rutine, det vi kan kalle prosjekt-dokumentasjon. Det er selvfølgelig ønskelig å se hele denne dokumentasjonen i sammenheng.

Et generelt krav til datadokumentasjonen bør være at den skal gi en total oversikt over alle tilgjengelige data med muligheter for lett søking etter klare og entydige beskrivelser av de data vi er interessert i til enhver tid. Systemet må samtidig gi den nødvendige beskyttelse mot illegal bruk av data. Dette gjelder særlig den tekniske beskrivelse av data i maskinlesbar form.

Generelt kan vi stille disse krav til dokumentasjon:

- Dokumentasjonen må være tilpasset de formål den skal tjene hos de grupper av fagfolk som skal bruke den: statistikkbrukere, statistikkprodusenter, de som utfører databearbeidingen. Dette leder til en spesialisering og inndeling av dokumentasjonen i tilsvarende grupper. For å lette samarbeidet mellom gruppene må de forskjellige deler av dokumentasjonen bindes sammen med presise referanser
- Dokumentasjonen bør, i størst mulig grad, framkomme som et ledd i arbeidet med planleggingen og gjennomføringen av et prosjekt, slik at en blir sikret en feilfri dokumentasjon
- Dokumentasjonen må føres best mulig à jour til enhver tid
- Det må være klare ansvarsforhold for utarbeiding og ajourhold av dokumentasjonen
- Dokumentasjonen bør utarbeides etter en vedtatt standard, som bør være godt beskrevet og enkel å lære
- Dokumentasjonen bør omfatte definisjon av enheter, bestand og kjennemerker, angi periode dataene refererer til og beskrive kvaliteten av dataene

1. Dokumentasjon beregnet på interne formål

Denne dokumentasjonen beskriver data som er hentet inn på spørreskjema, data på maskinlesbare media og publikasjoner.

a. Spørreskjema og skjemaarkiv ved Trykningskontoret

Trykningskontoret opprettholder et sentralt skjemaarkiv som omfatter alle spørreskjema som utarbeides av Byrået og alle andre institusjoners skjemaer som Byrået bearbeider. Rettledning for skjema-utfylling arkiveres sammen med skjema. Utfylte spørreskjema som Byrået låner av andre institusjoner, arkiveres i de tilfelle fagkontorene melder fra til Trykningskontoret. Spørreskjema som brukes ved andre institusjoner, arkiveres i den utstrekning slike skjema legges fram for Trykningskontoret for uttalelse.

Alle spørreskjema blir tildelt et nummer ved Trykningskontoret. Dette nummeret blir ikke byttet ut, selv om det blir foretatt endringer på nye utgaver av skjemaet. Nummer som blir ledige ved at skjema går ut av bruk, blir tatt i bruk til nye skjema. Et skjema er derfor identifisert ved hjelp av nummer og periode (årgang).

Arkivet er ordnet etter skjemanummer. For hvert skjema er det dessuten et kartotek kort med enkelte tekniske data. Disse kortene er ordnet etter kontor og skjemanavn.

Individualdata hentes i stor grad inn på spørreskjema som er utarbeidd av Byrået. Periode, enheter og kjennemerker i disse dataene må sies å være dokumentert ved teksten på spørreskjemaet og rettledningen for utfyllingen av skjemaet. Hvilken bestand det hentes inn opplysninger om, hvem som har gitt oppgavene og kvaliteten av opplysningene blir ikke dekket. Det vil heller ikke være praktisk å føre slike opplysninger på skjemaene. I noen tilfelle hentes oppgaver inn på magnetbånd eller hullkort. Det vil ikke alltid forekomme skjema for slike data.

1) Kapitlet bygger på et internt notat utarbeidd i forbindelse med Byråets langtidsprogram for 1976-80 (Notat nr. 8). Det arbeides nå med å forbedre Byråets dokumentasjonsrutiner.

Opgaver som behandles maskinelt, er dokumentert ved Systemkontorets filebeskrivelser (se pkt. c). I en del tilfelle arkiveres også spørreskjema sammen med rettledning for skjemautfylling og revisjons- og kodeinstruks i Systemkontorets prosjektdokumentasjon.

Skjemaarkivet har vært lite brukt utenfor Trykningskontoret og er heller ikke organisert med tanke på oppslag og søking etter informasjon.

b. Oversikt over Byråets sentrale arkiv for individualdata på spørreskjema

Så lenge det er behov for det, oppbevares utfylte spørreskjema ved fagkontorene. Deretter overføres skjemaene til Byråets sentrale arkiv, hvor de er ordnet etter statistikknavn og periode. Det er utarbeidd en spesiell arkivplan for arkivering og makulering av utfylte spørreskjema. Arkivplanen har ikke til hensikt å gi noen dokumentasjon av data, men viser til enhver tid hva som befinner seg i arkivet og hvor lenge det skal arkiveres.

Utfylte spørreskjema som ikke lenger er aktuelle for statistisk utnytting, overføres etter nærmere avtale til Riksarkivet.

c. Beskrivelse av data på maskinlesbare media

De fleste maskinlesbare data er lagret på magnetbånd. Dokumentasjonsteknikken for disse data har vært noe endret siden de første data ble lagret på bånd i 1962. De siste 7 årene har imidlertid teknikken vært den samme, og praktisk talt alle data i arkivet er beskrevet ved hjelp av den teknikk som brukes i dag. Teknikken kan også brukes for data på hullkort, men dette har i dag liten betydning.

Den største dataenheten i dokumentasjonssystemet er file. En file består av flere records. Litt forenklet kan en si at en record inneholder de kjennemerker for en enhet som det er praktisk å bearbeide samtidig. En file inneholder records (enheter) som bearbeides samtidig. "Samtidig" vil normalt si i samme program i datamaskinen.

F i l e k a t a l o g

Driftskontoret produserer maskinelt en katalog som viser alle filer og hvilken periode (årgang, kvartal osv.) av filen som nå står i arkivet. Filekatalogen produseres på grunnlag av filebeskrivelsene for produksjon.

For referanse begge veier mellom filekatalogen og filebeskrivelsene nyttes filenes identifikasjonsnummer. I tillegg nyttes statistikknummeret fordi filebeskrivelsene og filekatalogen av praktiske grunner er hovedsortert etter dette. Katalogen blir brukt til å vise hvilke data som finnes i arkivet og til markering av hvilke data som skal slettes.

F i l e b e s k r i v e l s e r f o r p r o g r a m m e r i n g

Filebeskrivelser for programmering er arkivert sentralt ved Systemkontoret. Arkivet er ordnet etter statistikknummer og er lett å bruke for oppslag. Ulempen ved arkivet kan sies å være at kjennmerkene i mange tilfelle er beskrevet ut fra et teknisk synspunkt, det er ikke lagt særlig vekt på at dokumentasjonen skal inneholde definisjoner for kjennmerkene. Fagkontorene har i svært liten grad deltatt i arbeidet med å dokumentere data på magnetbånd, det er også sjelden at filebeskrivelsene er blitt lagt fram for fagkontorene.

Filebeskrivelsene utarbeides under systemarbeidet og inneholder data om filen og hvert enkelt kjennemerke, men viser ikke hvilken periode eller generasjon av filer som er arkivert (dette går fram av filebeskrivelse for produksjon). Det er derfor ikke mulig fra denne dokumentasjonen å hente fram data fra båndarkivet uten å gå om filekatalogen eller filebeskrivelser for produksjon. Dette forhold er av sikkerhetsmessig betydning.

Kjennemerker som er overført fra skjema er i filebeskrivelsene bare beskrevet i den grad det er nødvendig for å utføre databearbeidingen. Dette har ført til at filebeskrivelsene ikke gir en dekkende beskrivelse av kjennmerkene. I de fleste tilfelle vil kjennmerkene være definert ved skjema, rettledning for utfylling av skjema og revisjons- og kodeinstruks.

Avledede kjennemerker er beskrevet i egne kodelister i prosjektdokumentasjon, i programbeskrivelsen eller de er innarbeidd i filebeskrivelsen. Kodelister som utelukkende er beskrevet i programbeskrivelsen er vanskelig tilgjengelige.

Kjennemerker som overføres fra andre datafiler er beskrevet etter en av de foran nevnte metoder.

Filebeskrivelse for produksjon

Dette er blanketter som føres manuelt ved Driftskontoret med utgangspunkt i en blankett som fylles ut av Systemkontoret når rutinen planlegges. Denne dokumentasjonen inneholder alle data fra filekatalogen, dessuten signatur for sletting av data.

Dokumentasjonen brukes til å rekvirere data fra arkivet og fylle ut kontrollkort for kjøring på datamaskinen. I kontrollkortet blir det satt inn en del data som maskinen kontrollerer mot tilsvarende data registrert på magnetbåndet. Dette sikrer oss at riktig data blir kjørt på riktig program.

Blanketten gir også plass for utfylling av filens omfang (bestand) og kjennemerker, men dette skal normalt av sikkerhetsmessige grunner ikke fylles ut. De første årene systemet var i bruk ble også disse data fylt ut.

Det har også her vært et problem å få plukket ut blanketter som etter hvert har blitt foreldet.

d. Sentrale kodelister

Kodelister er en del av datadokumentasjonen. For en del kjennemerker finnes det vedtatte standarder som blir brukt på tvers av statistikkområder. For de fleste kjennemerker er det ingen slike standarder. Byrået har ikke noe sentralt organ for registrering, identifisering og ajourhold av kodelister.

e. Prosjektdokumentasjon ved kontorene

Byrået har ingen fastlagt standard for dette, og praksis varierer fra kontor til kontor. Ved enkelte kontorer inngår bearbeidingsplaner, skjema og kodeinstrukser i projektdokumentasjonen.

Til prosjektdokumentasjonen bør det stilles disse utgangskrav¹⁾:

- (i) Dokumentasjonen må være fullstendig, det vil si at den må gi opplysninger om enhets- og kjennemerke-definisjon, om bestand og observasjonsperiode og om innsamlings-, kontroll-, bearbeidings-, publiserings- og lagringsrutiner for alle lagrede data
- (ii) Dokumentasjonen knyttes til det enkelte statistikk- eller analyseprosjekt. Ved prosjektavgrænsingen følges reglene i driftsstatistikken, og prosjektnummeret derfra nyttes også som identifikasjon av prosjektet i dokumentasjonssystemet
- (iii) Arbeidet med dokumentasjonen bør begynne i planleggingsfasen og videreføres ved fagkontor, system- og driftskontor etter hvert som arbeidet skrider fram. Den endelige sammenstilling og redigering av dokumentasjonen foretas av prosjektleder ved prosjektets fullføring
- (iv) Grunndokumentasjonen bør samles i ett notat for hvert prosjekt med tittelen Prosjektdokumentasjon for Det utarbeides regler for innhold og redigering mv. av notatene
- (v) For løpende og periodiske prosjekter som gjentas uten endringer i samsvar med tidligere dokumentert plan, utarbeides ikke nye dokumentasjonsnotat. Alle avvik fra tidligere beskrivelser av dataene dokumenteres imidlertid gjennom nye fullstendige dokumentasjonsnotater
- (vi) Fra hovedregelen om at grunndokumentasjonen samles i ett notat, bør det være unntak for statistiske standarder og maskinprogram. Som selvstendige deler av grunndokumentasjonen bør det opprettes ett arkiv for statistiske standarder og ett arkiv for maskinprogram
- (vii) I framstillingen av dokumentasjonsnotatene vil det være rasjonelt å ta i bruk elektronisk tekstbehandlingsutstyr. På den måten vil dokumentasjonen bli tilgjengelig både i manuelt lesbar og i maskinlesbar form

1) Hentet fra et prosjektnotat av I. Møglestue om system for datadokumentasjon.

f. Publikasjoner

Arkivet er et sentrallager for Byråets publikasjoner. Publikasjonene er ordnet etter publikasjonsserie og nummer. Et kartoteksystem som er utarbeidd for hver publikasjonsserie viser publikasjonens navn, når den er utgitt, tilgjengelige årganger og restbeholdning.

Kartotekene er ordnet for oppslag etter publikasjonens navn. I de tilfelle publikasjoner deles eller slås sammen er dette angitt på kartotekkortene.

g. Spørsmålsarkiv

Underavdelingen for intervjuundersøkelser har utarbeidd et spørsmålsarkiv. Arkivet er i første rekke et hjelpemiddel ved spørsmålskonstruksjon.

2. Dokumentasjon som dekker brukerne av statistikken

Denne delen av dokumentasjonen er i prinsippet et ekstrakt av dokumentasjonen som dekker interne formål og gir en kortfattet beskrivelse av de viktigste statistikkproduktene med referanser brukerne kan nytte for å skaffe seg flere opplysninger.

a. Veiviser i norsk statistikk

Veiviser i norsk statistikk gir en oversikt over statistikkområdene og hvilke publikasjoner som presenterer statistikkproduktene. Veiviseren refererer bare til publikasjonene. Disse refererer igjen til en viss grad til utrykte tabeller.

Veiviseren er systematisert etter emne etter samme standard som Årbokas kapittelinndeling. Den har også et alfabetisk stikkordregister. Ny utgave kommer første halvår 1980.

b. Hvem svarer på hva?

Hvem svarer på hva er en alfabetisk opplisting av forskjellige emneområder med referanse til medarbeidere i Byrådet. Oversikten utarbeides en gang pr. år. Den fordeles internt, sendes via F-reg til 2 600 av mottakerne i F-reg (bl.a. avisene) og til ca. 500 betalende abonnenter.

c. Publikasjoner fra Statistisk Sentralbyrå

Denne publikasjonslisten utgis en gang pr. år i to utgaver, en norsk og en engelsk. Den norske sendes til nær 3 000 av de faste mottakerne i F-reg og til ca. 1 000 betalende abonnenter. Den engelske utgaven sendes til ca. 500 utenlandske mottakere etter F-reg.

Publikasjonslisten har NOS-titlene i alfabetisk orden, og systematisert etter emne. Den inneholder også priser, opplysninger om hvor publikasjonene fås kjøpt og om den spesialservice Byrådet kan yte, gratis eller mot betaling.

d. Månedens publikasjoner

Denne listen blir sendt ut i begynnelsen av hver måned. Den gir en kort omtale av de publikasjoner som ble sendt ut i foregående måned, og en oversikt over publikasjoner som ventes i løpet av inneværende måned.

e. Publikasjoner

I tillegg til selve publikasjonene, som er den viktigste dokumentasjonsformen overfor utenforstående brukere, er det i enkelte publikasjoner referert til utrykte tabeller. Hittil er det bare referert til tabeller som er kjørt ut, men vil i framtiden også inkludere tabeller som er lette å produsere på standardprogram.

Utkommet i serien Statistisk Sentralbyrås Håndbøker (SSH)

- Nr. 1 Regler for publikasjonenes utstyr m.v. i serien Statistisk Sentralbyrås Håndbøker
- " 7 Prinsipper for konstruksjon av statistiske blanketter
- " 8 Framlegg til nordisk statistisk terminologi
- " 9 Standard for næringsgruppering ISBN 82-537-0906-4
- " 10 Hjemmel for innkreving av oppgaver
- " 12 Adresseliste for folkeregistrene
- " 13 Standard for handelsområder ISBN 82-537-0715-0
- " 22 Statistisk testing av hypoteser ved regresjonsberegninger
- " 23 Utsnitt om prinsipper og definisjoner i offisiell statistikk ISBN 82-537-0196-9
- " 24 Standard for gruppering av sykdommer - skader - dødsårsaker i offentlig norsk statistikk ISBN 82-537-0217-5
- " 25 Veiledning for brukere av den økonomiske modell MODIS III
- " 27 Utsnitt om prinsipper og definisjoner i offisiell statistikk. Fra Forbruksundersøkelsen 1958
- " 28 Standard for utdanningsgruppering i offentlig norsk statistikk ISBN 82-537-0272-8
- " 29 Norsk-Engelsk ordliste
- " 30 Lov, forskrifter og overenskomst om folkeregistrering ISBN 82-537-0099-7
- " 31 Håndbok for bruk av NATBLES ISBN 82-537-0100-4
- " 32 Konsumprisindeksen ISBN 82-537-0775-4
- " 33 Håndbok for bruk av DATSY ISBN 82-537-0329-5
- " 34 Sosialøkonomi. Kurs III på Byråskolen ISBN 82-537-0373-2
- " 35 Standard for kommuneklassifisering ISBN 82-537-0465-8
- " 36 Produksjonsindeks for bergverksdrift, industri og kraftforsyning ISBN 82-537-0947-1
- " 37 13 konjunkturindikatorer - En kort oversikt ISBN 82-537-0659-6
- " 38 Internasjonal standard for varegruppering i statistikken over utenrikshandelen (SITC-REV. 2) ISBN 82-537-0874-2
- " 39 Den statistiske behandlingen av oljevirkomheten ISBN 82-537-0702-9
- " 40 Regler for redigering og utstyr i publikasjoner fra Statistisk Sentralbyrå ISBN 82-537-0831-9
- " 41 Norsk i Byrået ISBN 82-537-0887-4
- " 42 Engrosprisstatistikk Engrosprisindeks Produsentprisindeks ISBN 82-537-0897-1
- " 43 Figurer i publikasjoner ISBN 82-537-1097-6
- " 44 Innføring i praktisk statistikk ISBN 82-537-1118-2

Pris kr 25,—

Publikasjonen utgis i kommisjon hos H. Aschehoug & Co. og
Universitetsforlaget, Oslo, og er til salgs hos alle bokhandlere.