

BRUK AV EN EDB-MASKIN I STATISTIKK- PRODUKSJONEN¹

Av byråsjef SVEIN NORDBOTTEN, Statistisk Sentralbyrå, Oslo

I. Oversikt

Fra å være et biprodukt av annen virksomhet har statistikkproduksjonen i et moderne samfunn utviklet seg til en stor og krevende databearbeidingsoppgave. Da Statistisk Sentralbyrå i Norge ble opprettet i 1876, ble dette arbeid tilfredsstillende utført av 15 funksjonærer. I U. S. A med sin raskt økende befolkning var en imidlertid allerede under bearbeidningen av folketellingen i 1880 bekymret for hvordan en skulle klare å bearbeide det store materialet som neste telling ville gi. Problemet ble løst av en funksjonær i det amerikanske statistiske byrå, Herman Hollerith, som konstruerte de første hullkortmaskiner.

Foregangsmannen i norsk statistikkproduksjon, direktør A. N. Kiær, gikk allerede i forbindelse med folketellingen 1900 inn for at en også i Norge skulle gjøre nytte av disse moderne hjelpemidler, og det er fra dette tidspunkt en kan regne den maskinelle databearbeiding i Statistisk Sentralbyrå.

Hullkortmaskinene er konstruert med sikte på å utføre bestemte operasjoner og er derfor et middel til rasjonalisering og automatisering av de enkelte bearbeidingsoperasjoner. Etter hvert som kravene til statistikken er blitt større og det kreves en mer intensiv utnyttning av det innsamlede materialet, har det oppstått behov for en integrasjon av bear-

¹ Artikkelen bygger på et foredrag som ble holdt i Stockholm den 25. mai 1959 etter innbydelse fra Matematikmaskinnämnden.

beidingsoperasjonene for å unngå at samordningsproblemet blir for stort. Dette førte til at det amerikanske Bureau of Census lot bygge den første elektroniske databearbeidingsmaskinen UNIVAC som det tok i bruk i 1952. I Norge ble en også klar over dette behov og studiene av EDB-systemer begynte her hos oss i 1952. I 1957 ble det i Byrået fullført systematiske studier og vurdering av de maskiner som ble bydd fram på det norske marked. Studiene viste at Byrået ville være tjent med en EDB-maskin, men at en i de første årene neppe ville makte å nytte kapasiteten fullt ut. Etter initiativ fra Statistisk Sentralbyrå ble det av Finansdepartementet oppnevnt et utvalg for å utrede spørsmålet om utnyttning av ledig kapasitet og endelig valg av maskintypen. Dette utvalg anbefalte i sin innstilling at det burde bevilges penger til kjøp av en EDB-maskin av typen DEUCE Mark II, og at Byrået skulle få leie ut ledig kapasitet til en selvstendig institusjon, Norsk Regnesentral, som også skulle fungere som en databearbeidingsentral for andre statsinstitusjoner, forskning og privat næringsliv.

I mai 1958 bevilget Det norske storting penger for kjøp av en EDB-maskin til Byrået og maskinen ble installert i løpet av vinteren 1958/59.

I den tid maskinen har vært i bruk, er den nyttet i bearbeidningen av vår månedlige statistikk for vareomsetningen med utlandet, ved en større forbruksundersøkelse, i bearbeiding av den årlige skat-

testatistikk og samferdselsstatistikk samt til statistiske beregninger. Vi planlegger å nytte maskinen i forbindelse med lønnsstillinger, industri- og varehandelsstatistikk og under bearbeidingen av folketellingen 1960.

Foruten de operasjoner som tidligere har vært utført maskinelt ved hjelp av hullkortmaskiner, har vi lagt stor vekt på å automatisere revisjonsarbeidet. Vi mener også at andre operasjoner som nå utføres manuelt med fordel kan automatiseres og smeltes sammen med andre operasjoner til en integrert prosess utført av DEUCE.

For å finne fram til en effektiv organisasjon for det forberedende arbeid som bruk av en EDB-maskin krever, har vi forsøkt flere opplegg og er nå kommet til at en sentralisering av dette arbeid hos en gruppe spesialister, fortrinnsvis rekruttert blant erfarne hullkortfolk, er den beste løsning.

Med den korte erfaring vi har, er det vanskelig å foreta noen vurdering av ulemper og fordeler ved bruk av en EDB-maskin i statistikkproduksjonen. Vi tror imidlertid nå at ulempene vesentlig består i det omhyggelige forarbeid som kreves og de mindre fleksible opplegg en derved får, mens fordelene er et større og kvalitetsmessig bedre statistikkprodukt enn det vi ellers ville kunne oppnå for samme kostnad.

2. Planer for omlegging av statistikkarbeidet

Umiddelbart etter avgjørelsen om valg av maskinen var tatt, gikk Byrået i gang med å forberede seg på omleggingen av enkelte statistikker til elektronisk databearbeiding. Det ble satt i gang såvel informasjons- som programmeringskurser i tilknytning til DEUCE i Byrået. Kursene med tilhørende håndbøker ble ut-

arbeidd av Byråets egen programmeringsgruppe som på denne tid hadde opparbeidd betydelig erfaring. Utover våren 1958, ble tallet på programmerere på full dag utvidet fra 2 til 4. Arbeidet med planlegging av maskinell databearbeiding ble utskilt i et eget kontor. Kontoret fikk som oppgave blant annet å stå for planleggingen og å koordinere alt maskinelt databearbeidingsarbeid i Byrået. Programmeringsarbeidet ble etter amerikansk mønster organisert i en programmeringsgruppe for hvert fagkontor. Hver gruppe besto av 3 funksjonærer hvorav en representerte fagkontoret, en hullkortavdelingen og en det nye planleggingskontoret. Det ble krevd at alle gruppemedlemmene skulle ha gjennomgått fullstendig treningskurser i programmering. For medlemmene av disse programmeringsgruppene ble det derfor holdt to 60-timers programmeringskurser i alt omkring 25 deltakere.

Arbeidet i 1958 omfattet i første rekke den månedlige handelsstatistikk og skattestatistikken. Samtidig blev bearbeidingen av den nye forbruksundersøkelsen 1958 allerede fra begynnelsen lagt opp med henblikk på DEUCE-bearbeiding. Dette valg av de første oppgaver ble dels bestemt av vårt ønske om å få de største databearbeidingsoppgavene snarest mulig over på elektronisk databearbeiding, samtidig som vi på grunn av vår ellers knappe maskinkapasitet ønsket å unngå nye og midlertidige utvidelser av vår hullkortpark.

Mens vår egen DEUCE var under installasjon, sendte vi folk til England for å teste programmene, idet vi tok sikte på å sette i gang produktive kjøringar på maskinen allerede fra første dag maskinen blev overtatt.

Maskinen blev overtatt den 10. mars i år, men da hadde vi allerede kjørt om-

kring 150 timer produktiv kjøring på den. Etter overtakelsen har utnyttelsen til ren produktiv kjøring av statistikkarbeider gjennomsnittlig vært 5—8 timer daglig.

I den planleggingsperiode som gikk foran overtakelsen av maskinen, gikk vi bevisst inn for en primitiv utnyttning av maskinen, dvs. i stor utstrekning å kopiere de konvensjonelle arbeidsrutiner. Begrunnelsen for dette var dels at vi da raskere ville kunne bruke maskinen og vinne erfaringer, og dels at vi også unngikk anskaffelse av nytt hullkortmaskinutstyr som ellers ville vært nødvendig fordi en del av vårt maskinanlegg var nedslitt.

3. Anvendelser av DEUCE i statistikkproduksjonen

3.1 Statistikkproduksjonens enkelte operasjoner

I Statistisk Sentralbyrå har vi funnet det hensiktsmessig å betrakte vår virksomhet som en produksjonsprosess. Dette har den fordel at en da kan dra nytte av det teoretiske begreppsapparat som er utviklet i tilknytning til bedriftøkonomikken.

Vi tenker oss i denne forbindelse produksjonprosessen delt opp funksjonelt i flere statistikkoperasjoner. Uten å gå nærmere inn på hva de enkelte betegnelser omfatter, regner vi med følgende standardoperasjoner:

- a. Administrasjon
- b. Planlegging og forarbeid
- c. Registerarbeid og oppgaveinnhenting
- d. Dataoverføring som omfatter punching
- e. Revisjon og koding
- f. Gruppering, beregning og tabellskriving

g. Kommentar- og analysearbeid

h. Publisering

i. Opplæring

De aller fleste av disse operasjoner omfatter rutinemessig databearbeiding, men det er særlig registerarbeid, oppgaveinnhenting, dataoverføring, gruppering og beregning som hittil har vært gjenstand for mekanisering og automatisering.

Et nytt felt hvor automatisering nå pågår vesentlig ved hjelp av DEUCE, er revisjonsarbeidet. Som jeg skal komme tilbake til senere, spiller automatiseringen av denne operasjon en meget stor rolle i vårt nåværende arbeid. Et annet felt som for andre enn oss i Norge kanskje synes å ligge utenfor rammen av statistikkproduksjonen, men hvor databearbeidingen er stor og åpenbart med fordel kan automatiseres, er analysearbeidet.

Den rekkefølge av statistikkoperasjoner som er nevnt ovenfor, var til dels tilfeldig bestemt av de konvensjonelle arbeidsrutiner en tidligere fulgte. Andre produksjonsteknikker betyr imidlertid en annen rekkefølge, oppspalting i deloperasjoner eller sammensmeltning av flere operasjoner til en integrert prosess. Etter hvert som de statistiske masser øker i omfang og samfunnets behov krever mer intensiv bearbeiding, vil videre integrasjon i automatiske prosesser være ønskelig for å redusere feilmuligheter og det økende administrative koordineringsproblemer.

3.2 Bruk av DEUCE i den månedlige handelsstatistikk

Vår største ordinære statistikk er den månedlige statistikk over vareomsættningen med utlandet. Hittil har denne statistikk utgjort mer enn 1/3 av Byråets databearbeidingskostnader. Oppgaven ble

tidligere kjørt på et eldre hullkortanlegg som Byrået eide, men som nå var nedkjørt. Statistikken omfatter oppgaver for opptil 130 000 innførte og 30 000 utførte varepartier pr. måned. Kort tid etter månedens utløp kreves foreløpige sumtall, og noen få uker deretter må de endelige tall foreligge. Det var således naturlig å ta fatt på denne arbeidsoppgave som en av de første.

I dag regner vi med 6 programmer for dette arbeid, hvorav 3 betraktes som ferdige mens det arbeides med de siste 3.

Det første gjelder kontroll og foreløpig summering av opplysningene til innførselsstatistikken. For hvert vareparti finnes det her opplysninger om partiets art, produksjonsland, innkjøpsland, tollsted, dets mengde og verdi. Opplysningene punches i et hullkort for hvert parti. Disse detaljkort sorteres deretter maskinelt i grupper sammen med spesielle masterkort som inneholder visse kontrollopplysninger og som senere gir grunnlag for kontroll av detaljkortene.

Denne sorterte kortmasse leses deretter inn i DEUCE som kontrollerer riktigheten av de opplysninger som finnes i detaljkortene, samtidig som den bygger opp de tabeller som er nødvendige. Maskinen skjelner mellom forskjellige typer feil og alt etter hvilken feiltype den finner, velger den selv etter programmet den spesielle behandlingsmåte som er relevant. Når DEUCE finner en gal opplysning i et detaljkort, punches den ut en kopi av kortet med indikasjon av feilens art. Disse såkalte feilkort blir manuelt gransket og eventuelt rettet, for deretter å bli sendt gjennom kontrollrutinen en gang til. Maskinen punches til slutt ut alle tabellene på sumkort som senere kan tabuleres på en hullkorttabulator.

Detaljene omkring de forskjellige be- tingede valg som maskinen foretar, er så

komplisert at det her vil føre for langt å gå inn på det. I stedet skal vi se på de forskjellige feil som kan forekomme og som maskinen kan påvise. Ved hjelp av masterkortene og spesielle tabeller som leses inn i maskinen sammen med programmet, kontrollerer maskinen hvert enkelt detaljkort etter samme prinsipper som revisorene tidligere gjennomgikk opplysningene. Maskinen går jo langt mer omhyggelig og systematisk til verks og kan påvise eventuelt:

- a. Urimelig høy enhetspris
- b. Urimelig lav enhetspris
- c. Urimelig høy verdi eller stor mengde
- d. Feil spesifisering av varens art
- e. Feil spesifisering av varens produksjonsland
- f. Urimelig kombinasjon av vareart og produksjonsland
- g. Feil spesifisering av innkjøpsland
- h. Feil spesifisering av tollsted

Enkelte av de ting maskinen påviser er direkte feil, mens andre ting er urimelig i forhold til en på forhånd fastsatt norm. Enkelte av de feilkort som punches ut fordi maskinen har påvist en urimelighet, kan derfor etter nærmere granskning vise seg å være korrekt. Slike feilkort vil uten videre kunne kjøres gjennom maskinen igjen og vil da bli automatisk godkjent og tatt med i tabelloppbyggingen.

Parallelt med denne kontrollen bygger DEUCE opp 4 forskjellige tabeller. 3 av disse har mengde og verdi i tabellhodet. I forspalten har den ene en hovedspesifisering av varens art med underspesifisering på produksjons- og innkjøpsland, mens den annen tabell foruten hovedspesifiseringen på varens art har en underspesifisering på tollsteder. Den siste tabellen gir bare en grov spesifisering av innførselen på varegrupper.

Foruten disse tabeller lager DEUCE

også en hjelpetabell som viser hvor mange feil av forskjellige typer som er funnet innenfor de forskjellige varegrupper under den maskinelle kontroll. Dette program omfatter således fragmenter både fra revisjonen og tabelleringsoperasjonene og er et steg mot integrasjon av disse operasjoner.

Et program etter liknende prinsipper er også utarbeidd for utførselen. For disse programmer som består av 6—800 ordrer er bearbeidingshastigheten omkring 6 000 detaljkort pr. time. Den viktigste bremse på denne hastighet er utpunching av sum- eller tabellkort for hver statistisk gruppe som skal spesifiseres i tabellene.

Mens vi kan si at hovedhensikten med de to nevnte programmer er å kontrollere og påvise feil, arbeides det også med to etterfølgende programmer hvis hovedformål er en videre reduksjon, dvs. kumulering i sumkort av materialet til andre statistiske tabeller, men hvor også kontrollen kommer med.

Foruten de tabeller som kontrollprogrammet produserer, kreves det også en rekke tabeller med kummulerte oppgaver for inn- og utførsel hittil i året. Disse sammendragsoppgavene skal i desember dessuten gi grunnlaget for årsstatistiken.

Reduksjonsprogrammet skal lese inn i sortert rekkefølge detaljkort som er kommet så sent at de ikke er kommet med i kontrollprogrammet, sumkort fra kontrollprogrammet og sammendragkort fra forrige måned. DEUCE skal først foreta en rutinekontroll av land- og tollstedskoder i alle kort og eventuelt påvise feil. Samtidig foretas en teknisk kontroll for å påvise dobbeltstansinger i en kolonne eller feil sortering. Maskinen undersøker deretter at de varer som skal ha dobbelt mengdeangivelse, f. eks. både

stk. og kg, har det og at det er rimelig forhold mellom de to angivelser. Videre undersøker den alle »brukte» varer. Pr. definisjon skal de nemlig ha lik kode for produksjons- og innkjøpsland. Ved uoverensstemmelser settes produksjonslandet lik innkjøpsland.

DEUCE tar så fatt på reduksjon av kortene til nye sammendragkort hvor det lages 4 forskjellige typer for hver statistisk gruppe. I det fjerde sammendragkort punches også gjennomsnittspris. Disse sammendragkortene identifiseres ved en 10-sifret kode og det gjøres markeringer for forskjellig bruk av kortene, f. eks. de som skal være med i månedsheftet og de som ikke skal være med der.

Programmet vil også samtidig med kontroll og reduksjon foreta visse korreksjoner av masterkortene som nyttes både i kontroll- og reduksjonsprogrammet. Ved endringer i en gjennomsnittspris vil f. eks. maskinen punche ut et nytt masterkort for grupper med tilsvarende korrigert toleranseområde for prisvariasjoner. Ved innlesning av spesielle korreksjonskort kan en dessutom foreta endringer i andre kontrollbetingelser.

Til slutt kan nevnes at vi skal utarbeide programmer for månedlige indeksberegninger for vareomsetningen med utlandet. Bearbeidingshastigheten for alle disse programmer vil bli bestemt av mengden på sammendragkort som punches ut.

Et morsomt eksempel på anvendelse av DEUCE for løsning av et publiseringsproblem holder vi i disse dager på å fullføre. Hos oss kan private firmaer abonnere på spesialopplysninger om enkelte varer. Tidligere ble disse opplysninger gitt ved fotostatkopiering av tabellister, men vi har funnet det mer effektivt å gå til anvendelse av DEUCE.

Rutinen er følgende: For hver vare har vi abonnementskort for alle som ønsker spesiaopplysninger om denne varen. Abonnementskortene er identifisert med et kundennummer. DEUCE produserer kopier av hvert varekort med de forskjellige kundennummer for hvert abonnementskort. Kopikortene for alle varenummer sorteres deretter på kundennummer, colleres sammen med ett navnekort for kunden, listes og legges i vinduskonvolutter. Bearbeidingskostnadene på DEUCE blir ca. 2 øre pr. vareabonnement.

For å gi et fullstendig bilde vil jeg også her nevne at vi har et spesielt kontroll- og reduksjonsprogram for utførsel av fisk. Dette programme blir kjørt hver uke og opplegget for det er i store trekk lik det som ligger til grunn for hovedprogrammene.

3. 3 Bruk av DEUCE i Forbruksundersøkelsen 1958

Allerede fra begynnelsen ble bearbeidningen av den nye forbruksundersøkelsen 1958 planlagt med sikte på bruk av DEUCE. Rutinen er også her lagt opp som 2 DEUCE-programmer, det første er et kontroll- og foreløpig reduksjonsprogram og dette er nå så godt som ferdig kjørt. Det annet er et reduksjonsprogram som ennå ikke er helt fastlagt. Foruten disse programmer er det også laget flere mindre hjelpeprogrammer, og vi håper at det ved hjelp av DEUCE også vil bli mulig å utføre regresjonsanalyser direkte på grunnmaterialet.

Materialet som omfatter omkring 3 600 husholdninger punchet på 3 600 masterkort og omkring 300 000 detaljkort. Masterkortene inneholder husholdningens fellesopplysninger mens hvert detaljkort opptil 5 vareposter.

Detaljkortene sammen med de tilhørende masterkort blev kjørt månedsvis

gjennom DEUCE. Kortene for hver måned var på forhånd sortert på husholdning ved masterkortet først. Maskinen kontrollerte selv at kortene kom i den riktige rekkefølge.

Som navnet tilsier hadde denne kjøringen på DEUCE som hovedoppgave å kontrollere de enkelte postene i kortet. Hver enkelt post måtte passere følgende kontroller:

- a. At varenummeret forekom i kodelisten
- b. At poster med mengde og verdi virkelig skulle ha oppgitt både mengde og verdi
- c. At poster med bare mengde eller bare verdi skal ha oppgitt bare mengde eller verdi
- d. At prisen for de postene som skal ha både mengde og verdi ligger mellom visse oppgitte prisgrenser
- e. At verdien for de postene som bare skal ha verdi ligger innenfor visse verdigrenser.

Poster som ikke passerte alle kontrollene på en tilfredsstillende måte ble ikke godkjent og det ble punchet ut et kopi av de detaljkortene som inneholdt en eller flere forkastede poster. For hver enkelt post fikk en dessuten punchet en markering som anga hvilken eller hvilke poster på kortet som var gale.

Ca. 30 varenummer i kodelisten var såkalte beholdningsvarer. For hver husholdning ble det punchet et kort pr. beholdningsvare. Disse kortene inneholdt sum mengde og verdi i 2 alternativer som vi var interessert i. Alternativ 1 var summene av mengde og verdi for de postene som var *innkjøpt* i løpet av måneden, mens alternativ 2 var summene av mengde og verdi for *forbruket* i måneden. Disse sammendragkortene inneholdt bare summen av godkjente

poster. Kortene måtte derfor korrigeres hvis det ble rettet på poster som hørte inn under beholdningsvarer.

Verdien av de godkjente postene ble addert opp i en tabell som omfattet 30 varegrupper. Summen for disse varegruppene ble punchet på kort, 5 varegrupper i hvert kort. Her var også beholdningsvarene inkludert etter alternativ 1 ovenfor. Da det bare var godkjente poster som kom med i summene, måtte disse sumkort rettes opp når feilen i detaljkortene var korrigeret.

For hver husholdning ble det dessuten punchet 2 avstemnings- eller kontrollkort. Det ene kortet, avstemningskortet inneholdt sum inntekter og sum utgifter. Det ble kontrollert at disse beløpene var like store. Det andre kortet inneholdt opplysninger om hvor mange poster som var godkjent, hvor mange som ikke var godkjent, hvor mange kort som inneholdt bare godkjente poster og hvor mange kort med en eller flere forkastede poster. Dette siste kortet er av spesiell verdi for senere vurdering av opplegget for bearbeidingen og de manuelle korreksjoner som er foretatt.

De 30 varegruppene blir nå tabulert etter alternative opplegg og gir grunnlag for både foreløpige tall og valg av de reduksjoner som en vil gjøre i neste del av bearbeidingen.

3. 4 Bruk av DEUCE i skattestatistikken

Statistisk Sentralbyrå utarbeider også en årlig skattestatistikk. Den er for personlige skattytere basert på et utvalg som omfatter omkring 200 000 personlige skattytere. Byrået får en vesentlig del av dette materialet på hullkort fra likningskontorene, og statistikken har i lengre tid vært hullkortbearbeidd. Vi fant nå at det ville være effektivt å bearbeide den ved hjelp av DEUCE.

Som for de andre oppgaver er også dette program lagt opp som et kombinert kontroll- og reduksjonsprogram. Maskinen leser inn detaljkort som på forhånd er sortert på fylke og skatteklasse, og kontrollerer blant annet:

- a. Formuesskatt til kommunene
- b. Inntektsskatt til kommunene
- c. Formuesskatt til stat
- d. Inntektsskatt til staten
- e. Alderstrygdavgift

Dersom maskinen finner uoverensstemmelser mellom det den beregner og det som allerede er punchet i kortet fra likningskontorene, markerer den feilen ved et feilkort som sendes tilbake til fagkontoret for nærmere gransking.

Maskinen foretar samtidig oppblåsning i henhold til de riktige utvalgsprosenter og oppbygning av tabellene. Hver gang den er ferdig med en statistisk gruppe som i dette tilfellet er en kombinasjon mellom fylke og skatteklasse, puncher den ut sumkort. Bearbeidingshastigheten for dette programmet er 6—7 000 kort pr. time.

Selv om en til neste år vil foreta visse modifikasjoner i programmet, har det nå vist seg å være meget effektivt og en god anvendelse av DEUCE.

3. 5 Bruk av DEUCE i kystfartsstatistikken

Som eksempel på anvendelse av DEUCE innen samferdselsstatistikken vil en her nevne et program for statistikken over løsfart på kysten. Dette er en årsstatistikk og kortmassen omfatter bare 15 000 kort pr. år.

Det lages et skipskort for hver båt og turkort for hver tur. Mens turkortene blir kontrollert på DEUCE blir skipskortene kontrollert på forhånd en gang for alle.

Kontrollen av turkortene omfatter i

hovedsaken indikasjoner og at utseilt distanse ligger innenfor det tillatte toleranseområde som gjelder for de forskjellige kombinasjoner av laste- og lossesteder. Feilkortene blir punchet ut med indikasjoner.

Maskinen utfører dessuten en ganske imponerende tabelloppbygning idet det i en gjennomkjøring bygges opp omkring 10 forskjellige og meget detaljerte tabeller. Likevel blir hele arbeidet nå fullført på et par timer, mens det ved hjelp av hullkortmaskiner tidligere krevde uker.

3.6 Anvendelser av DEUCE for regresjonsanalyser og kalkulering

En har pekt på at vi i Norge også inkluderer analysearbeid i statistikkproduksjonen. Ofte omfatter analyseberegningene store rutinemessige databearbeidingsoppgaver og det er derfor naturlig å prøve å nytte DEUCE også på dette felt. I motsetning til de oppgaver som det har vært pekt på ovenfor, hvor vi selv har måttet utføre all programmering, har vi for analyseberegningene et rikholdig programbibliotek til hjelp.

Eksempelvis vil jeg her få nevne regresjonsberegninger som vi har foretatt for vår Forskningsavdeling. Regresjonsprogrammet som har vært nyttet er meget fleksibelt og kan operere med opptil 30 variable simultant i alle retninger. Vi kan etter ønske få ut opptil 15 forskjellige resultat typer som blant annet inkluderer momentmatriser, inverterte momentmatriser, korrelasjonsmatriser, regresjonskoeffisienter, multiple regresjonskoeffisienter, standardavvik for regresjonskoeffisientene osv.

For beregninger av størrelseorden 6 variable anslo vi nylig at DEUCE på en time utførte et arbeid som svarte til minst 200 timeverk med elektriske bord-

maskiner. For større og mer kompliserte analyser blir forholdet selvsagt enda gunstigere.

En annen anvendelse som nylig ble foretatt for vårt nasjonalregnskapskontor var en enkel kalkuleringsoppgave. Den var imidlertid teknisk noe komplisert fordi en rekke faktorer på et kort skulle divideres med faktorer spredt på andre kort. Arbeidet har tidligere krevd mer enn 30 timer på hullkorttabulator. Med DEUCE ble det gjort på 3 kvarter. Den spesielle programmering som krevdes kostet ikke mer enn operatørtiden tidligere ved kalkulatoren.

4. Planlagte anvendelser av DEUCE

4.1 Lønnstillinger

Et av de felter hvor vi nå driver programmeringsarbeid er i tilknytning til lønnstatistikken. Mens kontrollarbeidene her består av kontroll av noen få koder og forholdstall, og derfor mest hensiktsmessig kan foretas ved hjelp av hullkortmaskiner, vil vi bruke DEUCE til selve tabellreduksjonen. For at vi imidlertid skal være helt sikre på at opplysningene har gjennomgått en manuell kontroll på forhånd, vil hvert kort med kontrollerte opplysninger få en spesiell markering. Vi vil senere la DEUCE kontrollere at denne markering finnes på kortet, og varsle dersom den finner kort uten en slik markering.

Alle tabellkort vil bli punchet ut i en gjennomkjøring. Kortene representerer delvis gjennomsnittstall og unntakelsesvis også prosenttall.

4.2 Industri- og varehandelsstatistikk

En annen av våre store årlige undersøkelser som nå programmeres er industristatistikken. Her kommer igjen kon-

trolloperasjonen i forgrunnen da oppgavene fra hver enkelt bedrift omfatter et meget stort antall kjennemerker som skal passe inn i et konsistent definisjonssystem, f. eks. bearbeidingsverdi skal være lik verdi av produksjon for egen regning pluss godtgjørelse for leiarbeid og reparasjonsarbeid minus verdi av råstoff og godtgjørelse for bortsatt arbeid. Simultant eksisterer det også et stort antall strukturrelasjoner mellom kjennemerkene, f. eks. mellom bearbeidingsverdi, sysselsetting, maskinutstyr og råstofforbruk.

Byrået nytter høyt kvalifiserte spesialister til å revidere materialet. Men ofte må en regne med et par års opplæringsperiode før revisorene er fullbafarne. I de senere år har vi tatt i bruk hullkortmaskinene til kontrollen og til dels også eksperimentert med mekanisk oppretting av de feil som ble påvist.

I år vil vi ta i bruk DEUCE ved kontroll og reduksjon og har store forventninger til denne anvendelse fordi vi her riktig kan dra nytte av maskinens egenskaper til en rekke simultane kontroller. På forhånd utarbeides toleranseområder for de forskjellige relasjoner idet en tar utgangspunkt i de erfaringer revisorene sitter inne med. Relasjonene beregnes under bearbeidningen for hver bedrift og sammenholdes så med toleranseområdet.

Som i våre andre anvendelser vil også tabellreduksjonen foregå samtidig med kontrollen. Det program vi vil nytte nå i år vil imidlertid være et rent provisorisk program idet rutineopplegget for industri- og varehandelsstatistikk trolig vil bli lagt om med det første. Det neste program vil ta hensyn til disse omlegginger og også gjøre en samkjøring av industri- og varehandelsstatistikken mulig.

Under dette statistikkområdet lages det også et spesielt program for vår kvartalsvise lagerstatistikk. Den kontroll som her skal utføres baseres på en vurdering av hva som er rimelige lagerendringer fra kvartal til kvartal for de forskjellige varer hos de forskjellige bedrifter.

4. 3 Folketellingen 1960

Vi er nå kommet til et punkt som trolig mange er spesielt interessert i å høre om, nemlig våre databearbeidingsplaner for folketellingen 1960. Dessverre må vi her med en gang si at vi ennå ikke har noen definitive planer idet vi først fra juli for alvor vil kunne konsentrere oss om denne oppgaven. Jeg vil derfor her avholde meg fra å gå inn på hvordan dette arbeid kan angripes men bare antyde et par synspunkter.

Vi har tidligere nærmest vært av den oppfatning at med det EDB-utstyr vi idag disponerer, vil store deler av kontrollarbeidet også denne gang mest effektivt utføres ved hjelp av en IBM 101, mens en del av tabellreduksjonsarbeidet kanskje kan utføres ved hjelp av DEUCE. Faren her vil trolig være at de forberedelser som kreves for å tilrettelegge materialet for DEUCE lett kan kreve mer enn en vinner ved bruk av den.

Folketellingen i Norge vil bli kombinert med en fiskeritelling og det som er sagt her gjelder også for den sistnevnte.

4. 4 Intern kostnadsanalyse

Som de fleste andre statistiske byråer strir også vi i Norge med å måle resultatene av vår virksomhet. Dette er jo selvsagt særlig viktig når en skal vurdere effektiviteten av alternative bearbeidingsformer. Hittil har vi ikke maktet å utvikle noen intern resultatsstatistikk, men vi har nå fått et meget brukbart opplegg for en kostnadsstatistikk med spesi-

fikasjon på statistikknummer, statistikkoperasjon, kostnadsart og kostnadssted. En del generalomkostninger som husleie, sentraladministrasjonen o. l. må fordeles på spesifikasjonene etter forskjellige prinsipper. Vi arbeider nå med et DEUCE-program som vi håper vil løse dette fordelingsproblem.

Foruten selve fordelingsproblemet planlegger vi også å inkludere tabellreduksjonen og analyserutiner for kostnadsstatistikken. Det er særlig viktig for Byråets ledelse å få en konsentrert og konsis oversikt over kostnadene og utviklingen presentert på en slik måte at oppmerksomheten rettes nettop mot de punkter i virksomheten hvor det skjer eller har skjedd noe.

Etter dette arbeid er det nærliggende å ta fatt på vårt lønnsregnskap som omfatter omkring 250 funksjonærer i et komplisert system bestående av lønnsklasser, alderstillegg, fradrags- og skattetrekkregler.

4. 5 Optimal allokering i utvalg

Resultatene fra DEUCE-kjøringen av kostnadsstatistikk vil forhåpentligvis bli en hjelp under planlegging og budsjettering av nye undersøkelser, men vi vil også anvende DEUCE i annet planleggingsarbeid, f. eks. ved opplegg av utvalsundersøkelser hvor en støter på en rekke optimaliseringsproblemer.

Ved opplegg av en slik undersøkelse har en ofte relevante observasjoner av enhetene i massen og oppgaven blir gjerne å velge det av et gitt sett av alternative utvalgsopplegg som gir maksimum av informasjon pr. krone. Allerede ved meget innsnevrede problemstillinger får en svært store sett av alternativer og om alle skulle studeres koster det meget tid og arbeid. Vi mener at det her må kunne utvikles et DEUCE-program for

de mest alminnelige problemstillinger hvor maskinen arbeider seg gjennom alle alternativer og registrerer den optimale løsning eller eventuelt ordner alle alternativer etter graden av optimalitet på grunnlag av på forhånd oppsatte kriterier.

4. 6 Løsning av visse statistiske fordelingsproblemer ved simulering

Et annet problem innen samme felt som vi planlegger å angripe ved hjelp av DEUCE er studier av visse statistiske fordelinger ved hjelp av simulering. Følgende problem interesserer f. eks. våre sampling-eksperter: Hvilke egenskaper kjennetegner fordelingen for en konsumprisindeks som bygger på estimater for en rekke priser på grunnlag av ett utvalg og konsumgrunnlag eller vektor estimert på opplysninger fra annet utvalg. Fordelingen for en slik indeks som er komponert av stokastiske variable med kjent fordeling hvis parametere kan estimeres, kan en ved hjelp av simulering utvikle og studere. Analoge simuleringsproblemer innen naturvitenskapen har for lengst vært løst ved hjelp av elektroniske regnemaskiner og vi håper også at vi skal kunne løse våre problemer ved hjelp av DEUCE.

5. Organisasjonsproblemer

5. 1 EDB-arbeidets plass i virksomheten og dets organisasjon

Maskinell databehandling i sin alminnelighet og EDB i særdeleshet får en stadig bredere plass i statistikkproduksjonen. I Statistisk Sentralbyrå anslo vi nylig at vel 70 prosent av vårt databehandlingsarbeid nå utføres rent maskinelt, og det er rimelig at en vier de organisasjonsmessige problemer dette reiser en tilsvarende interesse.

Det amerikanske statistiske byrå, Bureau of the Census, anbefaler lokal programmering av databearbeidingen. Vi startet også med en slik organisasjonsform, men den virket ikke tilfredsstillende hos oss. Årsaken er vel hovedsakelig våre mindre forhold som ikke gjør det mulig å holde spesialister i databearbeiding i fullt arbeid på hvert fagkontor. Vi har derfor sentralisert databearbeidingen. Dette er forøvrig i overensstemmelse med den organisasjonsmessige utvikling ellers idet vi i den senere tid har hatt en utvikling fra organisasjon etter statistikkområder til organisasjon etter operasjoner.

Planleggingskontoret for maskinell databearbeiding som skal planlegge, vurdere forskjellige bearbeidingsformer og bruken av vår maskinkapasitet, utarbeider i samarbeid med fagkontorene kjøreplaner og programmer mens ansvaret for gjennomføring og betjening av maskinen er pålagt vår maskinavdeling. Begrunnelsene for å skille planlegging fra drift er mange. Jeg vil her bare nevne at erfaringene viser at dersom spesialister ikke blir isolert i egne organisasjonsmessige »båser», er det alt for fristende å bruke dem til løpende arbeid når det kniper. Planleggingskontoret har også som oppgave å følge med i den tekniske utvikling på databearbeidingsområdet, slik at Byrået til enhver tid har muligheten for å tilpasse seg de tekniske hjelpemidler som finnes.

La oss se litt nærmere på arbeidet som utføres ved planleggingskontoret. EDB-arbeidet omfatter hos oss analyse og programmering av arbeidsoppgavene for elektronisk bearbeiding, prøving av programmene, kjøring av maskinen, utdanning av programmerere og operatører, studier og vurdering av elektronisk utstyr. Av størst interesse her vil

trolig være programmeringen, dvs. tilretteleggingen av problemene for løsning på EDB-maskinen.

Det første trinn i programmeringen er utarbeiding av *hoveddiagrammet* som skal gi en oversikt over hele databearbeidingsoppgaven hvor anvendelse av DEUCE inngår på ett eller flere steder. For hver anvendelse av DEUCE lages deretter et *funksjonsdiagram* med sikte på å få fram de enkelte funksjoner, f. eks. spesielle kontroller, beregninger og reduksjoner som er lagt over på DEUCE. I tilknytning til disse diagrammer hører også en metode- eller systembeskrivelse. Programmeringens to første trinn har derfor et analytisk preg og kan ved databearbeidingsoppgaver kreve relativt lang tid. Ved matematisk pregede oppgaver vil disse deler av programmeringen ofte være gitt alt ved problembeskrivelsen.

Det tredje trinn er utarbeidingen av det *logiske diagram* hvor funksjonene spaltes opp og organiseres i overensstemmelse med EDB-maskinens spesielle arbeidsteknikk. Her begynner arbeidet å få sin spesielle tekniske karakter som særlig blir framtrædende i det såkalte *flow-diagram* som er det fjerde trinn. I flow-diagrammet uttrykkes arbeid ved de spesielle elementæroperasjoner DEUCE kan utføre. Lagerplassen fordeles, tiden beregnes og de enkelte deler av programmet justeres og tilpasses hverandre. Det femte og siste trinn i programmeringen er *koding* av flow-diagrammer etter en tallkode som maskinen kan tyde. Arbeidet krever stor nøyaktighet men byr ellers ikke på store problemer. Under visse betingelser kan en overlate maskinen selv å utføre kodingen.

Programmeringen krever folk med spesielle anlegg og utdanning. Det må imidlertid her nevnes at det er utviklet for-

skjellige typer lettkodingsystemer som gjør DEUCE tilgjengelig for en videre krets av brukere.

Til programmeringen hører også omhyggelig prøvekjøring av programmene. Dette er et tidkrevende arbeid, men legger, når det blir planmessig utført, heldigvis relativt lite beslag på maskinene. Produsenten av DEUCE og en sammenslutning av brukerne, DEUCE Users' Association, har samlet en rekke programmer for standardproblemer i et programbibliotek. For den type databearbeidingsoppgaver som Byrået nå har, må vi imidlertid utføre de fleste programmene selv.

5. 2 Rekruttering og opplæring av EDB-personale

Alle kjenner trolig påstanden om at det er lettere å utdanne en statistiker til programmerer enn omvendt. Vi arrangerte derfor fullstendige programmeringskurser for et, etter våre forhold, stort antall funksjonærer fra forskjellige kontorer og med forskjellig bakgrunn. Et av formålene var å lokalisere de funksjonærer eller grupper av funksjonærer som syntes å ha anlegg for programmeringsarbeid. Resultatet vi kom fram til og som vi nå har basert oss på, er at databearbeidingen i statistikkproduksjonen med det maskinanlegg vi i dag har vil kreve at programmererne er fullstendig fortrolig med hullkortarbeid og opplegg, de må ha lang erfaring i statistikkbearbeiding, de må ha utpregede evner for logisk tenkning, være utholdende og gjerne ha matematisk skoleing. For DEUCE, som er en binær-maskin, er spesielt litt kjennskap til tall-systemer en fordel.

Vi har rekruttert våre programmerere blant hullkortavdelings aller beste folk. Et par har vi sendt på 3-ukers kurser

hos produsenten, men nå mener vi at vi selv like effektivt kan stå for opplæringen. Hvor lang tid tar det så før en programmerer er vel kvalifisert? For folk som tilfredsstiller de kvalifikasjoner som blev antydnet ovenfor, mener vi at etter et halvt års programmering under rettleiing og kontroll av erfarne programmerere skulle opplæringsperioden kunne anses som avsluttet og de nye programmerere arbeide selvstendig. Dette kan kanskje for mange synes å være en lang tid, men så må er også huske på at det er et stort arbeidsområde som strekker seg fra analyse av arbeidsoppgaven til skriving av operatørinstrukser som programmererne våre får ansvar for.

Om kort tid vil Statistisk Sentralbyrå ha 6 utdannede programmerere på hel dag. I dag mener vi at dette skulle svare til vårt behov for omlegging til maskinell bearbeiding og til opplegg for nye arbeidere i de nærmeste årene framover.

6. Fordeler som er oppnådd ved EDB

Hvilke fordeler har vi så hittil oppnådd ved bruk av DEUCE? Etter bare få måneders bruk har vi et svakt grunnlag for å uttale oss, men vi kan likevel si: Hittil har automatiseringen av databearbeidingen først og fremst gitt seg utslag i mer pålitelig bearbeiding, muligheter for behandling av større masser og mer utnyttning av det innsamlede materialet. Foreløpig er det en mekanisering av de konvensjonelle rutiner som er foretatt. Etter hvert må vi imidlertid gå inn for en integrasjon av operasjonene som også trolig vil gi oss samordningsmessige fordeler og raskere bearbeiding. Når det gjelder statistikkens aktualitet må en imidlertid huske at den er avhengig av den seneste operasjonen

i kjeden, og den hører ikke alltid med under bearbeidingen.

For kort tid siden foretok vi en undersøkelse ved de fagkontorer som nå har de største arbeider på DEUCE for å høre hva de mente om den elektroniske databearbeidingen.

Kontoret for handelsstatistikken mente at det uten den maskinelle kontroll ved hjelp av DEUCE, ville vært umulig å oppnå den kvalitet statistikken i dag har. Overgangen til ny varenomenklatur har ført til en stor øking i varespesifikasjonen. Hvilke konsekvenser dette ville hatt uten DEUCE er vanskelig å si. Kontoret kan derfor ikke selv peke på noen kostnadsreduksjoner, men mener at det er gode muligheter når systemet blir mer innarbeidd. Når det gjelder maskinbearbeidingen kan en imidlertid vise til en besvarelse på netto kr. 50 000 pr. år i innspart hullkortmaskinleie.

Kontoret som leder Forbruksundersøkelsen 1958 sier at samme kvalitet ville det vært praktisk ugjørlig å oppnå uten DEUCE. Den eneste bremsen på bearbeidingen har foreløpig vært det manuelle arbeid som må gjøres ved kontoret. På første del av undersøkelsen ble det brukt mellom 30 og 40 DEUCE-timer. En liknende bearbeiding på vanlige hullkortmaskiner ville krevd punching av mer enn 4 ganger så mange hullkort som vi i dag nytter og nesten uoverkommelig kalkulering og tabulering som også ville krevd nye maskiner. De programmeringskostnader vi har hatt på denne statistikken svarer omtrent til de operatørkostnader vi ellers ville fått.

Også for skattestatistikken hevdes det at kvaliteten er blitt langt bedre enn tidligere, samtidig som maskinkostnadene er redusert med omkring 50 prosent.

De økonomiske fordeler vi oppnår ved

DEUCE er selvsagt avhengig av driftssikkerhet og avskrivningstakt. Hittil har vi, i de vel 2 måneder vi har nyttet maskinen, ikke hatt noe å klage på når det gjelder driftssikkerheten av maskinen. Dette til tross for at maskinen til sine tider har vært i bruk opptil 20 timer i døgnet.

Etter mange overveielser ble det funnet realistisk å sette avskrivningsperioden til 8 år. Takket være en utleieavtale med Norsk Regnesentral og med en egen utnytting av 1 500 timer på maskinen vil vi i år internt få en meget lav og fordelaktig timepris for bruk av DEUCE. Dette gjør det økonomisk forsvarlig for oss å nytte maskinen på en rekke felter som vanligvis faller utenfor EDB-maskinenes anvendelse. Programmeringskostnadene vil variere sterkt fra oppdrag til oppdrag. Våre egne programmer omfatter gjennomsnittlig 500 ordrer og krever gj. 5 mnd. programmering.

7. Framtidig utvikling av databearbeidingen

Hittil har vi arbeidd med automatisering av de enkelte operasjoner, særlig av kontroll og tabellreduksjoner som til en viss grad er integrert. Vi har også nevnt noen eksempler på hvordan DEUCE er nyttet til analyseformål og hvordan vi håper å nytte den også i planleggingsfasen.

Etter hvert som samfunnets behov for statistikk krever mer intensiv bearbeiding av det statistiske materiale, vil vi få et stort administrativt samordningsproblem. Nettopp på dette felt tror vi at den elektroniske databearbeiding særlig vil vise seg å bli verdifull fordi rutinemessige samordningsproblemer i stor utstrekning kan formuleres som databearbeidingsoppgaver.

Den ideelle løsning vil da bli en integrert automatisk statistikkproduksjon hvor EDB-anlegget programmeres til selv å samordne det innsamlede materiale, sine enkelte deler og operasjoner ved bearbeidingen fram til det ferdige analyserte sluttprodukt presentert i den ønskede form.

En slik integrert bearbeiding innebærer imidlertid mange uløste problemer og krever et meget utbygd EDB-system. De kan bare løses gradvis over en lengre perioden.

Først har vi revisjon- og kodeoperasjonen. Foreløpig har vi bare løst en del av denne, nemlig kontroll av oppgavene med påvisning av eventuelle feil. Kodeproblemet vil trolig kunne løses i stor utstrekning ved å nytte prekodning under skjemakonstruksjonen. Oppretting av feil er derimot langt vanskeligere. I U. S. A. har en i lengre tid nyttet meget enkle regler for automatisk oppretting, men våre forløpige eksperimenter tyder på at for intensiv bearbeiding av våre små masser fører de amerikanske prinsipper galt avsted. Trolig må vår løsning være av følgende type: Påviste feil grupperes i 2 klasser, nemlig alvorlige og uvesentlige feil alt etter hvorledes de kan påvirke de statistiske sluttresultater. De uvesentlige feil rettes automatisk i EDB-systemet etter bestemte regler mens maskinen for alvorlige feil skriver ut et relevant revisjonsskriv adressert til oppgavegiveren. Svaret må senere automatisk kunne innarbeides i de endelige resultater.

Neste trinn vil bli å smelte analysen inn i den integrerte prosess. Her vil vi prøve å få andre karakteristika ved massen, først og fremst spredningsmål som er av almen interesse men særlig også er avgjørende for statistisk planlegging av nye undersøkelser. Disse beregninger

vil trolig uten større merkostnader kunne tas med i de fleste programmer.

Et annet interessant spørsmål som det nå arbeides med i Byrået, er utarbeiding av bestemte tabeller ved hjelp av regresjonsanalyse i stedet for simple anslag på kjennemerkene i de enkelte ruter av tabellen. De arbeider som er gjort hittil synes å være meget lovende, samtidig som arbeidet uten større problemer kan integreres med andre deler av databearbeidingen.

Dette fører oss over til presentasjon eller publisering av sluttproduktet. Allede i mange år har vi preget stensiler til noen av våre stensilerte publikasjoner direkte ved hjelp av hullkorttabulator, og vi har også foretatt direkte tabulering av manuskripter for publikasjoner som skulle offset-trykkes. Med moderne skriveutstyr i tilknytning til EDB-systemet skulle vi kunne utvikle dette enda lenger samtidig som det nå på grunn av lang trykkesetid er meget som tyder på at stensilpublikasjonen kan bli en mer alminnelig form for publisering av de statistiske resultater. Vi har også tenkt oss mulighetene av at EDB-systemet automatisk utformer et styreprogram for en typografisk settemaskin.

Går vi tilbake i statistikkproduksjonen igjen, møter vi den arbeidskrevende dataoverføringsoperasjonen som i dag utføres ved hjelp av punchmaskiner. Alle er trolig kjent med de eksperimenter som pågår i U. S. A. om automatisk dataoverføring av den fotoelektroniske FOSDIC. Anvendelse av slikt utstyr vil stille visse krav til skjemakonstruksjonen av samme art som prøvekodningen og det kan kanskje også bli spørsmål om en annen innsamlingsteknikk.

Til slutt vil det oppstå et integrasjonsbehov på tvers av statistikkene for å nytte ut alle de kombinasjonsgruppering-

er som her foreligger, f. eks. kryssgruppering på grunnlag av materiale både fra industristatistikk og finansstatistikk. Dette vil trolig etterhvert føre til produksjon av mer »arkivstatistikk», dvs. statistikkproduksjon på grunnlag av arkivopplysninger uavhengig av innsamlingen som bare utføres for å holde ajour arkivene. Slik statistikkproduksjon vil kanskje gjøre det mulig å oppnå en mer strømlinjeformet virksomhet uten toppbelastning i enkelte perioder.

Vi har nå pekt på noen av de problemer og fordeler en integrert automatisk bearbeiding vil by på, og det kan være grunn til også å understreke at den har visse ulemper. Først og fremst krever den en lang planleggingsperiode med relativt meget større investeringer i det forberedende arbeid enn vi hittil har vært vant til. En slik statistikkproduk-

sjon vil derfor bli lite fleksibel og vanskelig å tilpasse til stadig varierende behov.

8. Sluttmerknad

Det er et optimistisk syn på mulighetene og utviklingen som har preget arbeidet i Statistisk Sentralbyrå, men en skal heller ikke legge skjul på at vi mange ganger er blitt skuffet i våre forventninger.

Likevel tror jeg at vi i dag er kommet lenger enn vi våget å håpe da vi startet arbeidet for 5—6 år tilbake. Vi har derfor ovenfor drøftet tildels dristige ideer og framtidige muligheter. Skal en makte å følge med i den rivende tekniske utvikling og dra nytte av den på sitt eget arbeidsfelt, er forutsetningen at en selv tør la tankene følge nye og dristige baner.