



*Katharina Henriksen*

# Notater

**Utvalgsplan til konsumpris-  
indeksens nye matvareindeks -  
Basert på strekkodedata**

# Innhold

1. Innledning .....	3
2. Ny indeks for matvarer og alkoholfrie drikkevarer i KPI .....	3
Bruk av strekkodedata gir nye muligheter i utvalgstrekingen.....	4
3. Trekkpopulasjonen og opprinnelig butikkutvalg .....	5
3.1. Populasjonen av skannerbutikker utgjør trekkpopulasjon.....	5
3.2. Opprinnelig butikkutvalg som leverer strekkodedata .....	6
4. Opplegg og gjennomføring .....	7
4.1 Stratifisering av trekkpopulasjon .....	7
4.2 Allokering av trekkpopulasjonen .....	12
4.3 Trekking av butikkutvalget .....	15
5. Valg av metode .....	17
6. Oppsummering.....	18
Referanseliste .....	21
De sist utgitte publikasjonene i serien Notater.....	22



## **1. Innledning**

I august 2005 publiserte Statistisk sentralbyrå (SSB) en forbedret delindeks for matvarer og alkoholfrie drikkevarer i konsumprisindeksen (KPI). Denne indeksen baseres utelukkende på strekkodedata fra de største aktørene innenfor dagligvarehandelen i Norge, samt fra en sentral aktør innenfor servicehandelen. Strekkodedataene er innhentet fra aktørenes hovedkontorer, og inneholder informasjon om blant annet omsatt mengde og gjennomsnittspris i en gitt periode for samtlige varer i utvalgsbutikkene.

Innhenting av elektroniske data og anvendelse av strekkodedata åpner for å etablere butikkutvalget til KPI på nye måter. Dette notatet dokumenterer prinsippene som er lagt til grunn og vurderinger som er gjort i forbindelse med trekkingen av et nytt butikkutvalg til den strekkodebaserte indeksen.

I kapittel 2 gis en beskrivelse av den nye delindeksen for matvarer og alkoholfrie drikkevarer.

KPI har i flere år mottatt strekkodedata, men har hittil kun benyttet prisinformasjonen for utvalgte produkter. I kapittel 3 gis en deskriptiv beskrivelse av KPIs opprinnelige utvalg av butikker som har levert strekkodedata og av hele butikkpopulasjonen for næringen 52.11 (SN02) basert på SSBs Bedrifts- og foretaksregister (BoF).

Kapittel 4 viser flere måter å organisere trekkpopulasjonen på som har vært vurdert for den nye utvalgsplanen. Det er vanligvis hensiktsmessig å inndele trekkpopulasjonen i mindre delpopulasjoner, kalt stratum. Kapitlet omtaler også hvordan total utvalgsstørrelse allokeres ut på hvert enkelt stratum, samt hvordan butikker bør trekkes fra hvert enkelt stratum.

Kapittel 5 redegjør for vårt valg av opplegg og metode for trekkingen av det nye butikkutvalget.

Kapittel 6 gir en oppsummering.

## **2. Ny indeks for matvarer og alkoholfrie drikkevarer i KPI**

Formålet til KPI er å måle den faktiske prisutviklingen for varer og tjenester som etterspørres av private husholdninger i Norge. Matvarer og alkoholfrie drikkevarer utgjør vektmessig omlag 12 prosent av KPI. Den tidligere delindeksen for matvarer og alkoholfrie drikkevarer i KPI benyttet en varekurv på om lag 250 representantvarer for å måle prisutviklingen. Prisene på disse varene ble innhentet fra et utvalg med 115 butikker. Disse butikkene ble trukket fra en bedriftspopulasjon basert på næringsplassering og en geografisk inndeling i åtte områder. Se Bråten (1996) og Statistisk sentralbyrå (2001) for ytterligere informasjon.

I den nye delindeksen for matvarer og alkoholfrie drikkevarer inngår over 14 000 varer hver måned. Strekkodedataene er elektroniske data som inneholder opplysninger om omsatt mengde og oppnådd gjennomsnittspris på et produkt i en bestemt butikk i et spesifisert tidsrom. Både den kraftige økningen i antall varer og det økte informasjonstilfanget knyttet til hver enkelt omsatte vare høyner kvaliteten på indeksen. Det beregnes flere mikroindekser enn tidligere, og opplysningene om omsatt mengde gjør det mulig å fange opp skift i etterspørselen etter varer. Se Haraldsen og Rodriguez (2005) for ytterligere dokumentasjon om metoder, egenskaper og resultater angående den nye delindeksen for matvarer og alkoholfrie drikkevarer.

Strekkodedata innhentes via hovedkontorene til de største aktørene innen dagligvarehandel (NorgesGruppen ASA, ICA Norge AS, Coop Norge AS og Rema 1000 Norge AS) og en sentral aktør fra servicehandelen (Narvesen Norge AS). Disse aktørene klassifiseres innunder næring 52.11 (Butikkhandel med bredt vareutvalg med hovedvekt på nærings- og nytelsesmidler) i BoF. I den nye delindeksen for matvarer og alkoholfrie drikkevarer inngår ikke butikker fra næring 52.20 (butikkhandel med nærings- og nytelsesmidler i spesialforretninger) som det gjorde tidligere.

## **Bruk av strekkodedata gir nye muligheter i utvalgstrekkningen**

Metoden som ligger til grunn for å trekke ut butikkutvalg i KPI er et resultat av avveininger mellom teoretiske hensyn og noen praktiske begrensninger. Ved å endre rutinene for datainnsamlingen for mat- og alkoholfrie drikkevarer til en sentralisert elektronisk løsning fjernes eller reduseres noen av de praktiske begrensningene som har vært forbundet med datainnhenting fra papirskjema i enkeltbutikker.

### **Butikkutvalget kan økes uten særlige økonomiske kostnader**

Størrelsen på butikkutvalget i KPI skal sikre at alle varer og tjenester er godt representert i undersøkelsen. Utvalgsstørrelsen er samtidig bestemt ut fra kostnader som påløper ved en undersøkelse. Det kan være kostnader i forbindelse med intervjubesøk og kostnader som påløper ved å håndtere ett butikkutvalg (vedlikeholde utvalget, utsendelse av skjema, mottak av skjema etc.). Ved bruk av strekkodedata faller mange av disse kostnadene bort. SSB trenger ikke lengre å gjennomføre førstegangsintervju for uttrukne butikker til undersøkelsen. Kostnader ved skjema utsendelse bortfaller også.

### **Bruk av intervjuere oppheves ved strekkodedata**

SSB benytter i stor grad besøksintervju for butikker som er trukket ut til å inngå i KPI. I disse intervjuene blir butikkene informert om undersøkelsen, samt at intervjueren bistår butikken med å velge ett representativt vareutvalg til undersøkelsen som butikken skal følge over tid. SSBs intervjukorps dekker ikke alle kommuner. Dette førte tidligere til at butikker i disse kommunene ble utelatt fra undersøkelsene i KPI. Disse butikkene utgjorde i underkant av 40 prosent av alle butikkene i næring

52.11. Disse butikkene i distriktene har forholdsvis lav omsetning på grunn av deres beliggenhet og kundemasse, slik at dette ikke har vært ett særlig stort problem for KPI. Intervjukorpset har særlig lav dekning av kommuner i Sogn- og Fjordane, Møre og Romsdal, Nordland og Finnmark. Ved å innhente strekkodedata fra kjedenes hovedkontor kan butikker i alle kommuner trekkes ut til undersøkelsen, såfremt de rapporterer elektronisk til deres hovedkontor.

### **Minimal oppgavebyrde for den enkelte butikk**

Ved bruk av strekkodedata reduseres oppgavebyrden for butikkene betraktelig ved at hovedkontoret har overtatt rapporteringsplikten. Kjedenes hovedkontor har i stor grad etablert automatiserte rutiner for overlevering av strekkodedata til KPI, slik at deres oppgavebyrde er minimal. Tidligere ble butikkpopulasjonen i næring 52.11 innskrenket ved at KPIs butikkutvalg ikke skulle overlappe SSBs undersøkelse av detaljomsetningen (DOI) sitt butikkutvalg. Dette for å unngå den belastende oppgavebyrden ved å rapportere til begge undersøkelsene. Dette medførte at om lag ti prosent av butikkene ble unnlatt fra KPIs trekkpopulasjon. Samtidig var det etablert en rutine at DOI trakk først ut butikkene i næringen basert på omsetningen. Dette resulterte i at KPI mistet mange av de største i sin trekkpopulasjon. I de senere årene har ikke KPI tatt hensyn til dobbelrapporteringen siden prisinnhenting har vært elektronisk.

## **3. Trekkpopulasjonen og opprinnelig butikkutvalg**

### **3.1. Populasjonen av skannerbutikker utgjør trekkpopulasjon**

Til tross for at spesialbutikker (bensinstasjoner, bakeriutsalg og spesialbutikker for kjøtt og fisk) selger matvarer og alkoholfrie drikkevarer har vi valgt å holde disse utenfor den nye delindeksen.

Trekkpopulasjonen for delindeksen utgjør dermed butikker som har elektronisk rapportering til hovedkontorene til følgende kjeder:

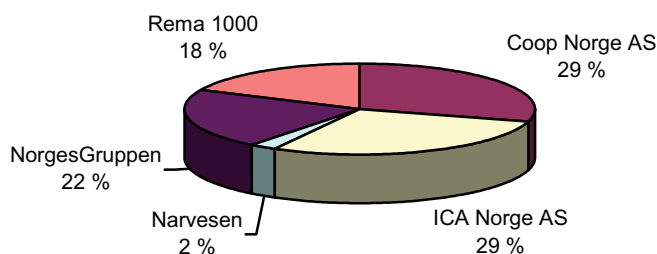
- NorgesGruppen ASA
- ICA Norge AS
- Coop Norge AS
- Rema 1000 Norge AS
- Narvesen Norge AS

BoF viser at vel 45 prosent av alle butikkene og godt over 70 prosent av omsetningen innen næring 52.11 er tilknyttet kjedene Narvesen, Rema 1000, ICA Norge, NorgesGruppen eller Coop Norge, og teller 3078 butikker. Disse 3078 butikkene utgjør *trekkpopulasjonen*, det vil si den populasjonen som vi trekker det nye butikkutvalget ifra. Butikker er *rapporteringssenheten* i undersøkelsen, mens matvarer og alkoholfrie drikkevarer er *observasjonsenheten* vi ønsker å måle prisutviklingen for.

Samlet årsomsetning er på 71,7 milliarder kroner, mens gjennomsnittsomsetningen er 23,3 millioner kroner. Ikke alle butikkene står oppført med omsetning eller opplysninger om hvilket år omsetningen gjelder i BoF. Hovedvekten av butikkene står oppført med 2002 som nyeste omsetningsår. Vi har valgt å inkludere butikker som står oppført med null eller ”missing” i omsetningen i trekkpopulasjonen dersom de er etablert etter 2002. Dette for å sikre oss om at nyetablerte butikker kan trekkes til butikkutvalget.

Det er verdt å merke seg at omsetningsbegrepet i BoF ikke samsvarer helt med butikkenes omsetning fra salg av matvarer og alkoholfrie drikkevarer. Dette gjør at visse kjeder kan bli noe overrepresentert i utvalget.

**Figur 1: Kjedenes andel av samlet årsomsetning på 71,7 milliarder kroner**



Kilde: Bedrifts- og foretaksregisteret

Figur 1 viser hvordan årsomsetningen på 71,7 milliarder kroner registrert i BoF fordeles på de ulike kjedene. Coop Norge og ICA Norge er de to kjedene med høyest omsetning, med omsetningsandeler på 29 prosent. NorgesGruppen og Rema1000 har noe lavere omsetningsandeler, mens Narvesen har en omsetningsandel på to prosent.<sup>1</sup> Salg av matvarer utgjør en minimal andel av Narvesens totale salg, slik at en omsetningsandel på to prosent i dette tilfellet antakelig vil være for høyt. Narvesen kan dermed bli noe overrepresentert i utvalget.

### 3.2. Opprinnelig butikkutvalg som leverer strekkodedata

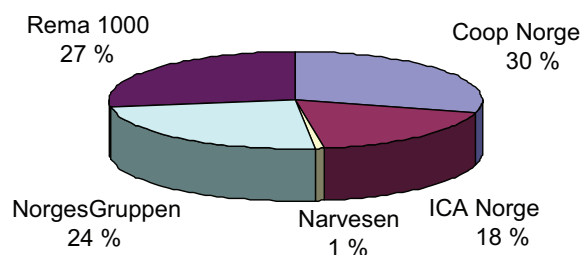
Det opprinnelige butikkutvalget i næring 52.11 ble trukket etter de samme prinsippene som gjelder KPIs hovedundersøkelse som følger SSBs utvalgsplan. Det vil si at trekkpopulasjonen stratifiseres etter geografisk beliggenhet. Der Norge er inndelt i fem geografiske områder (Akershus, Østlandet med unntak av Oslo, Sørlandet - Vestlandet med unntak av Bergen, Møre - Trøndelag med unntak av

<sup>1</sup> Ifølge ACNielsens Dagligvarefasit for 2002 var dagligvarebransjens totale netto omsetning på 92,2 milliarder kroner. Bedrifts- og foretaksregisteret viser at total omsetning for næring 52.11 er 99,9 milliarder kroner i 2002. Se <http://www.dagligvarehandelen.com>

Trondheim og Nord-Norge), og deretter er Oslo, Trondheim og Bergen inndelt som egne soner. Videre trekkes butikkutvalget etter PPS-prinsippet ( Probability Proportional to Size sampling).

Med denne fremgangsmåten utgjorde utvalget i KPI fra næring 52.11 115 butikker som rapporterte elektronisk. Disse butikkene hadde en samlet årsomsetning på 5,4 milliarder kroner.

**Figur 2: Kjedenes omsetningsandeler i eksisterende butikkutvalget som leverer strekkodedata utgjør 5,4 milliarder kroner**



Kilde: Konsumprisindeksen

Figur 2 viser KPIs utvalg av butikker som leverer strekkodedata innen næring 52.11 fordelt på kjedene. I utvalget er særlig Rema 1000 overrepresentert i forhold til trekkpopulasjonen (se figur 1).

NorgesGruppen er noe overrepresentert, mens ICA Norge og Narvesen er underrepresentert.

Omsetningsandelen til Coop Norge er om lag like store i populasjonen og utvalget.

## 4. Opplegg og gjennomføring

### 4.1 Stratifisering av trekkpopulasjon

Å inndelegge trekkpopulasjonen inn i disjunkte delpopulasjoner, kalt strata, er en hensiktsmessig måte å organisere trekkpopulasjonen på, og er mye brukt i statistiske utvalgsplaner. Hensiktene med stratifisering er å garantere en viss utvalgsstørrelse i hvert enkelt stratum samt redusere usikkerheten i estimatene for totalomsetningen.

#### Måter å stratifisere

Det er mange måter å stratifisere trekkpopulasjonen. Det er ønskelig å lage de enkelte strata slik at de er mest mulig homogene med tanke på prisutvikling.



I utgangspunktet ønsket vi å stratifisere trekkpopulasjonen inn i følgende strata:

- Geografisk område (Akershus, Oslo, Resten av Østlandet, Sørlandet - Vestlandet, Bergen, Møre - Trøndelag, Trondheim og Nord-Norge)
- Kjede (NorgesGruppen, ICA Norge, Rema 1000, Coop Norge og Narvesen)
- Kjedenes konsept (Kiosk, Lavpris, Nærmarked og Supermarked)

Det gamle butikkutvalget i delindeksen for matvarer og alkoholfrie drikkevarer var stratifisert etter geografisk områder. Vurderingen med å stratifisere etter geografisk område er basert på ønsket om å sikre oss at disse handelsdistriktene blir tilstrekkelig representert i butikkutvalget. Stratifisering etter geografisk område, kjeder og konsept vil imidlertid medføre at butikkutvalget må bli ufordelaktig stort dersom vi vil ha en rimelig utvalgsstørrelse i hvert stratum.

Alle kjedene i trekkpopulasjonen er viktige aktører innen detaljhandelen i Norge, og vi ønsker derfor at det nye utvalget skal være representert med butikker fra alle disse kjedene. En høy grad av kjederepresentasjon i utvalget bidrar til å sikre at KPI følger prisutviklingen på varer som er viktige for husholdningene.

De fleste aktørene i detaljhandelen opptrer med flere ulike konsepter innad i deres konsern. Faktorer som pris, varesortiment og beliggenhet synes å ha en viktig betydning i deres utforminger av konseptprofiler. Disse faktorene er vektlagt hos kjedene for å gi de enkelte konseptene ulike særpreget og for at dekke kundenes ulike behov.

Ved en gjennomgang av aktørenes konseptinndeling har vi kommet frem til en inndeling i fire ulike konsepter (kiosk, lavpris, supermarked og nærbutikk). Noen kjeder opererer med færre eller flere konsepter enn de fire valgte. Rema1000 butikkene vil kun inngå i lavpriskonseptet. NorgesGruppen har også butikker som profilerer seg som store supermarked og som stormarkeder, mens Coop Norge har butikker som betegnes hypermarkeder. På tross av at disse butikkene har et større vareutvalg og trolig enn annen prisprofil har vi valgt å la disse inngå i stratumet supermarked.

Med vår valgte stratuminndeling inndeles butikkene i Coop Norge som følger:

- Kiosk: Strax-kiosk
- Lavpris: Prix
- Supermarked: Mega og Obs
- Nærmarked: Marked og S- Nærkjøp

Følgende inndeling får butikkene i ICA Norge:

- Lavpris: Rimi
- Supermarked: Ica Supermarked, Fokus og Ica Maxi
- Nærmarked: Ica Sparmat, Livi og Servicemat

Følgende inndeling får butikkene i NorgesGruppen:

- Kiosk: Mix
- Lavpris: Kiwi, Bunnpris
- Supermarked: Centra, Spar, Ultra, Meny og Eurospar
- Nærmarked: Joker og Nærmat

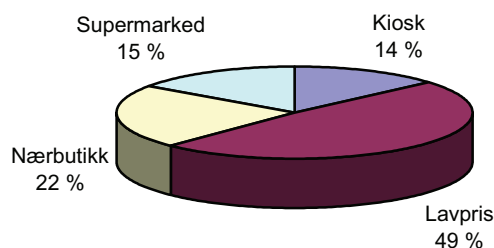
Denne inndelingen i kjedetilknypning og konseptprofil gir tretten ulike strata. Tabell 1 gir en opptelling av antall butikker i hvert enkelt stratum. Vi ser av tabellen at kiosksegmentet hos Coop Norge (COOP) og NorgesGruppen (NG) har få butikker.

**Tabell 1: Stratuminnndeling basert på kjedenes konsept**

Stratum nr	Konsept	Antall butikker	Prosent
1	COOP KIOSK	12	0,4
2	COOP LAVPRIS	329	10,7
3	COOP NÆRMARKED	370	12,0
4	COOP SUPERMARKED	181	5,9
5	ICA LAVPRIS	551	17,9
6	ICA NÆRMARKED	286	9,3
7	ICA SUPERMARKED	75	2,4
8	NARVESEN	371	12,1
9	NG KIOSK	38	1,2
10	NG LAVPRIS	301	9,8
11	NG NÆRMARKED	35	1,1
12	NG SUPERMARKED	213	6,9
13	REMA 1000	316	10,3
		3078	100,0

Det er etablert ett arbeid mellom seksjon for økonomiske indikatorer, seksjon for bedriftsregister og seksjon for bygg- og tjenestestatistikk for å bedre oversikten av butikkstrukturen i BoF. En bedre identifisering av butikkenes kjede- og konserntilknytning øker kvaliteten i butikkutvalget ved at hver butikk er sikret riktig konserntilknytning, kjedetilknypning og konsept. Sett opp mot nyere tall fra Dagligvarehandelens oversikt over hovedkontor og kjeder i Norden avviker vårt antall butikker til dels mye fra deres opplysninger. Det er særlig for butikker innen NorgesGruppen som er underrepresentert i vår trekkpopulasjonen. Etableringen av en egen kjedetilknypning i BoF vil forhåpentlig redusere dette avviket i fremtiden.

**Figur 3: Fordelingen av antall butikker i de ulike konseptene**



Fra figur 3 kan vi se at trekkpopulasjonen til den nye delindeksen for matvarer og alkoholfrie drikkevarer er sterkt preget av butikker som tilhører lavpriskonseptet. Butikkene i lavprissegmentet fremstår som et konsept for hverdagsinnkjøp med et begrenset vareutvalg og relativt lave priser. Varesortimentet er i stor grad basert på basisvarer. Nærbutikkkonseptet synes å fokusere mer på personlig service og ha en viss grad av lokale tilpasninger i varesortiment og pris. I motsetning til lavprisbutikkene profileres supermarkedene som et konsept med større mangfold av varer, ett høyere servicenivå, samt ett høyere fokus på ferskvarer. Det er ikke etablert egne stratum for større supermarkeder/hypermarkeder, og slike butikker inngår dermed i stratumet for supermarked. Det bør vurderes å ha egne strata for disse største supermarkedene på sikt.

#### **Populasjonsparametre i hvert enkelt stratum**

I utvalgsplanen til det nye butikkutvalget er omsetningen valgt som størrelsesmålet trekningen skal gjøres etter. Omsetning brukes ofte som proxyvariabel for å følge prisutviklingen, og benyttes også i KPIs hovedundersøkelse. Omsetningsvariabelen hentes fra BoF, og er innregistrert via SSBs næringsundersøkelse. Av praktiske årsaker er omsetningen for flerbedriftsforetak i BoF minst to år forsinket.

Tabell 2 viser hvordan omsetningen til det totale antall butikker fordeles på hvert enkelt stratum. Totalomsetningen for hele populasjonen utgjør 71,7 milliarder kroner, mens gjennomsnittlig omsetning er 23,3 millioner kroner.

**Tabell 2: Inndeling i stratum etter konsept. Populasjonsparametre, 1000 kroner.**

Stratum	Konsept	Antall butikker	Totalomsetning	Gjennomsnitt	Standardavvik
1	COOP KIOSK	12	44721	3727	3896
2	COOP LAVPRIS	329	6227203	19043	9607
3	COOP NÆRMARKED	370	3787235	10236	9854
4	COOP SUPERMARKED	181	11036621	61315	64667
5	ICA LAVPRIS	551	15078999	28291	14216
6	ICA NÆRMARKED	286	2371748	8351	6303
7	ICA SUPERMARKED	75	3109000	46403	57066
8	NARVESEN	371	1710345	4972	3663
9	NG KIOSK	38	138584	4470	7338
10	NG LAVPRIS	301	6016997	22040	11085
11	NG NÆRMARKED	35	187256	6041	3536
12	NG SUPERMARKED	213	9057054	44397	32388
13	REMA 1000	316	12967153	43808	20113
		3078	71732916	23305	

Ikke uventet har supermarkedskonseptet høy gjennomsnittlig omsetning per butikk. Særlig supermarked i Coop Norge har i snitt høy omsetning per butikk, men også supermarkedene innen NorgesGruppen og ICA Norge har en betydelig omsetning per butikk.

Kjedenes supermarkedskonsept har en større spredning, målt ved standardavviket, i omsetningen enn de andre konseptene. Dette tyder på at kjedenes supermarkedskonsept har en mindre homogen butikkstruktur enn hva som gjelder i de andre strata. Dersom vi hadde skilt ut de større supermarkedene i egne strata ville sannsynligvis spredningen i omsetningen blitt betraktelig mindre i kjedenes supermarkedskonsept. Eksempelvis profilerer ICA Norge deres MAXI butikker som stormarkeder hvor forbrukerne ved siden av dagligvarer kan kjøpe bøker, husholdningsartikler, klær og sportsartikler. Coop Norges tilsvarende stormarkedsprofil er OBS butikkene. NorgesGruppens MENY butikker og ULTRA butikker representerer deres profil for et stort vareutvalg og stort fokus på ferskvarer.

Innen lavpriskonseptet dominerer Rema1000, men også de andre kjedene har i snitt høy omsetning per butikk. Vi ser av tabellen at spredningen for kjedenes lavpriskonsept er mye mindre enn deres supermarkedsprofil. Gjennomsnittlig omsetning er lavest for kjedenes konsepter for lokale butikker (nærbutikker) og kiosksalg. Innen nærbutikkkonseptet er Coop Norge størst i omsetning per butikk, mens ICA Norge og NorgesGruppens nærbutikker innbringer i snitt noe lavere omsetning. Narvesen som har en ren kiosksprofil har en betydelig større totalomsetning enn hos Coop Norge og NorgesGruppen, men i snitt er omsetningen kun noe større per butikk.

## 4.2 Allokering av trekkpopulasjonen

Nå skal vi se nærmere på hvordan den totale utvalgsstørrelsen kan allokere ut på hvert enkelt stratum.

### Total utvalgsstørrelse på 150 butikker

Vi har på forhånd bestemt at den totale utvalgsstørrelsen bør være omlag 150 butikker. Dette mener vi er en hensiktsmessig størrelse på et butikkutvalg. Dette er en utvalgsstørrelse vi kan håndtere effektivt, blant annet ved at det eksisterer noen tekniske begrensninger ved å overføre mye større datafiler.

### Måter å allokere på

I arbeidet med ny utvalgsplan for trekking av butikkutvalget for delindeksen for matvarer og alkoholfrie drikkevarer ble særlig to ulike måter vurdert;

- proporsjonal allokering
- optimal allokering (Neyman)

*Proporsjonal allokering* innebærer at den totale utvalgsstørrelsen fordeles ut på hvert stratum etter samme forhold som populasjonen er fordelt på strata. Med andre ord:

$$n_l = n \left( \frac{N_l}{N} \right) = n W_l \quad l = 1, \dots, L$$

hvor  $n_l$  er allokert utvalgsstørrelse i stratum  $l$ ,  $n$  er total utvalgsstørrelse,  $N_l$  er populasjonsstørrelsen i stratum  $l$  og  $N$  den totale populasjonsstørrelsen. Dermed blir  $W_l$  populasjonsandelen. Vi benytter både antall butikker og omsetning som allokeringensvariabel.

*Neyman allokering* gir en optimal måte å allokere på når man tar hensyn til stratas størrelse og deres interne variasjon i omsetningen

$$n_l = n \frac{W_l S_l}{\sum_{k=1}^L W_k S_k}$$

hvor  $n_l$  er allokert utvalgsstørrelse i stratum  $l$ ,  $n$  er total utvalgsstørrelse,  $W_l$  er populasjonsandelen i stratum  $l$  og  $S_l$  er populasjonsstandardavviket i stratum  $l$ .  $\sum_{k=1}^L W_k S_k$  er produktet av populasjonsandelen og standardavviket, summert over alle strata. Formelen viser at den totale utvalgsstørrelsen bør fordeles slik at man får store utvalg i strata som er store og hvor populasjonsvariansen er stor. Populasjonsstandardavvikene  $S_l$  må være kjent på forhånd i alle strata for at formelen kan benyttes. Vi vet også at dersom spredningen er like innen alle strata kan  $S_l$  forkortes bort i formelen. Da vil vi stå tilbake med formelen for proporsjonal allokering.

### **Effekten av proporsjonal allokering**

Proporsjonal allokering medfører at utvalgsstørrelsen i hvert enkelt stratum gjenspeiler trekkpopulasjonens størrelse innenfor de ulike strata. Tabell 3 oppsummerer hvordan den totale utvalgsstørrelsen på 150 butikker fordeler seg på hvert enkelt stratum når allokeringen skjer proporsjonalt etter henholdsvis antall butikker og omsetningen.

**Tabell 3: Allokering av total utvalgstørrelse på hvert enkelt stratum, proporsjonalt med antall butikker og proporsjonalt med omsetning, 1000 kroner**

Stratum	Konsept	Antall butikker		Proporsjonal allokering etter antall butikker			Proporsjonal allokering etter omsetning		
		Total	omsetning	Stratums- andel	Utvalgs- størrelse	Utvalgs- andel	Stratums- andel	Utvalgs- størrelse	Utvalgs- andel
1	COOP KIOSK	12	44721	0,004	2	0,17	0,001	2	0,17
2	COOP LAVPRIS	329	6227203	0,107	16	0,05	0,087	13	0,04
3	COOP NÆRMARKED	370	3787235	0,120	18	0,05	0,053	8	0,02
4	COOP SUPERMARKED	181	11036621	0,059	9	0,05	0,154	23	0,13
5	ICA NORGE LAVPRIS	551	15078999	0,179	27	0,05	0,210	32	0,06
6	ICA NORGE NÆRMARKED	286	2371748	0,093	14	0,05	0,033	5	0,02
7	ICA NORGE SUPERMARKED	75	3109000	0,024	4	0,05	0,043	7	0,09
8	NARVESEN	371	1710345	0,120	18	0,05	0,024	4	0,01
9	NG KIOSK	38	138584	0,012	2	0,05	0,002	2	0,05
10	NG LAVPRIS	301	6016997	0,098	15	0,05	0,084	13	0,04
11	NG NÆRMARKED	35	187256	0,011	2	0,06	0,003	2	0,06
12	NG SUPERMARKED	213	9057054	0,069	10	0,05	0,126	19	0,09
13	REMA 1000	316	12967153	0,103	15	0,05	0,181	27	0,09
		3078	71732916	1,000	152		1,000	157	

Effekten av proporsjonal allokering etter antall butikker er at den totale utvalgstørrelsen på 150 butikker fordeles på hvert enkelt stratum tilsvarende hvordan butikkantallet i trekkpopulasjonen er fordelt på de ulike strata. Ved proporsjonal allokering basert på antall butikker i trekkpopulasjonen er utvalgsandelen  $150/3078=0,05$ . Vi ser av tabellen at stratumet med Coops kiosksalg og NorgesGruppens nærbutikker er noe overrepresentert i forhold til trekkpopulasjonen (0,17 og 0,06). Dette skyldes ønsket vårt om at alle strata skal være representert med minst to butikker i det nye butikkutvalget. Vi har derfor satt krav i forbindelse med allokeringen at alle strata skal inneholde minst to butikker.

En proporsjonal allokering etter omsetning gir i utgangspunktet ingen butikker til stratum 1 (Coop kiosk), 9 (NorgesGruppen kiosk) og 11 (NorgesGruppen Nær). Også her ønskes minst to butikker per strata. Vi ser av tabellen at proporsjonal allokering etter omsetning gir en litt annen fordeling enn en allokering etter butikkantallet. De strata som inneholder supermarkeder blir tildelt flere butikker, mens de fleste andre strata får en lavere utvalgsandel ved proporsjonal allokering etter omsetning. I dette tilfellet tildeles også Rema1000 og lavpriskonseptet til ICA Norge flere butikker. Allokeringsmåten påvirker særlig Narvesen. Deres kiosksalg er mange, men med relativ lav omsetning. Dette gjør at de ble tildelt fjorten færre butikker når allokeringsvariabelen er omsetning enn ved en proporsjonal allokering etter butikkantallet.

## Effekten av Neyman allokering

Ved en Neyman allokering tas det hensyn til spredningen innen hvert enkelt stratum i trekkpopulasjonen når fordelingen av butikker skjer i tillegg til antall butikker. Strata som består av mange butikker og hvor spredningen i omsetningen er høy bør tildeles relativt flere butikker enn små strata med liten spredning. En Neyman allokering vil dermed redusere usikkerheten i estimatet for totalomsetning. Vi ser av tabell 4 at det tildeles forholdsvis mange butikker til supermarkeds-konseptene. Dette er middels store strata, men på grunn av stor spredning i omsetningen tildeles de forholdsvis mange butikker. Lavpriskonseptet til ICA Norge tildeles også forholdsvis mange butikker. Dette skyldes først og fremst størrelsen på stratumet, og ikke på grunn av spredningen i omsetningen.

**Tabell 4: Allokering av total utvalgsstørrelse på 150 butikker på de enkelte strata ved Neymans metode, 1000 kroner**

Stratum Konsept	Antall butikker	Total omsetning	Gjennom snitt	Standard avvik	Stratumsandel	Utvalg andel
1 COOP KIOSK	12	44721	3727	3896	0,0039	2 0,16667
2 COOP LAVPRIS	329	6227203	19043	9607	0,10689	9 0,02736
3 COOP NÆR	370	3787235	10236	9854	0,12021	11 0,02973
4 COOP SUPERMARKED	181	11036621	61315	64667	0,0588	35 0,19337
5 ICA NORGE LAVPRIS	551	15078999	28291	14216	0,17901	23 0,04174
6 ICA NORGE NÆR	286	2371748	8351	6303	0,09292	5 0,01748
7 ICA NORGE SUPERMARKED	75	3109000	46403	57066	0,02437	13 0,17333
8 NARVESEN	371	1710345	4972	3663	0,12053	4 0,01078
9 NG KIOSK	38	138584	4470	7338	0,01235	2 0,05263
10 NG LAVPRIS	301	6016997	22040	11085	0,09779	10 0,03322
11 NG NÆR	35	187256	6041	3536	0,01137	2 0,05714
12 NG SUPERMARKED	213	9057054	44397	32388	0,0692	20 0,0939
13 REMA 1000	316	12967153	43808	20113	0,10266	19 0,06013

155

## 4.3 Trekking av butikkutvalget

Hittil har vi sett på flere hensiktsmessige måter å organisere trekkpopulasjonen – stratifisering av kjedenes konsepter og hvordan utvalgsstørrelsen kan allokere etter ulike kriterier. Neste steg er å bestemme hvordan vi skal trekke de enkelte butikkene som skal inngå i den nye delindeksen for matvarer og alkoholfrie drikkevarer.

Det er ønskelig å benytte ett opplegg som sikrer at hver enkelt butikk har en kjent sannsynlighet for å bli uttrukket til det nye butikkutvalget. Det finnes mange alternative trekkemetoder, men vi skal se på *enkel tilfeldig trekking uten tilbakelegging* og *pps-trekking*. Begge disse metodene gir et sannsynlighetsutvalg.



Ved *enkel tilfeldig trekking uten tilbakelegging* fås et utvalg der hver butikk innen hvert enkelt stratum har lik sannsynlighet for å bli trukket til å inngå i butikkutvalget.

Ved en *pps- trekking* er butikkenes sannsynlighet til å bli trukket til butikkutvalget proporsjonalt med butikkens omsetningsstørrelse.

I et enkelt tilfeldig utvalg er inklusjonssannsynligheten  $\pi = \frac{n}{N}$  lik for alle butikkene i trekkpopulasjonen,  $n$  betegner total utvalgsstørrelse og  $N$  er totalt antall butikker i trekkpopulasjonen.

Dersom vi trekker ett stratifisert butikkutvalg med *enkel tilfeldig trekking uten tilbakelegging* når trekkpopulasjonen er *allokert etter antall butikker* vil hver butikk innen hvert enkelt stratum få en inklusjonssannsynlighet lik  $150/3078 = 0,05$ . Vi ser av tabell 3 at inklusjonssannsynligheten er lik utvalgsandelen i hvert enkelt stratum. Vi husker at stratumet med Coops kiosksalg og NorgesGruppens Nærbutikker i utgangspunktet inneholdt for få butikker til å bli tildelt noen butikker. Med tilleggskriteriet om at minst to butikker skal allokeres til hvert enkelt stratum gjør at disse strata får en noe høyere inklusjonssannsynlighet.

Dersom vi trekker ett stratifisert butikkutvalg med *enkel tilfeldig trekking uten tilbakelegging* når trekkpopulasjonen er *allokert etter omsetning* vil fortsatt hver enkelt butikk innen hvert enkelt stratum ha lik sannsynlighet for å bli trukket ut. Inklusjonssannsynligheten i de ulike strata vil derimot variere siden utvalgsstørrelsen er allokert etter omsetning – se tabell 3. De strata som har butikker med relativ høy omsetning vil bli tildelt flere butikker enn stratum bestående av butikker med lavere omsetning.

Dersom allokeringen er gjort etter *Neymans* metode er det vanlig å benytte *enkel tilfeldig trekking uten tilbakelegging* som trekkemetode. Dette vil sikre en optimal allokering. Ved Neymans metode tas butikkenes spredning i omsetningen med som ett kriterium i fordelingen av butikkene på hvert enkelt stratum. I tillegg skal store strata med hensyn på antall butikker tildeles flere butikker enn mindre strata.

Dersom *pps - trekking* skal benyttes er det mest naturlig at total utvalgsstørrelse blir *allokert etter antall butikker*. Det er mulig å allokere proporsjonal med omsetning, og deretter trekke proporsjonalt med butikkenes omsetningsstørrelse, men både allokeringmetoden og trekkemetoden gir en oversampling av butikker med høy omsetning. Kombinert vil de gi en ytterligere oversampling av butikker med høy omsetning, og det er derfor vanlig å nøye seg med en av delene.

## 5. Valg av metode

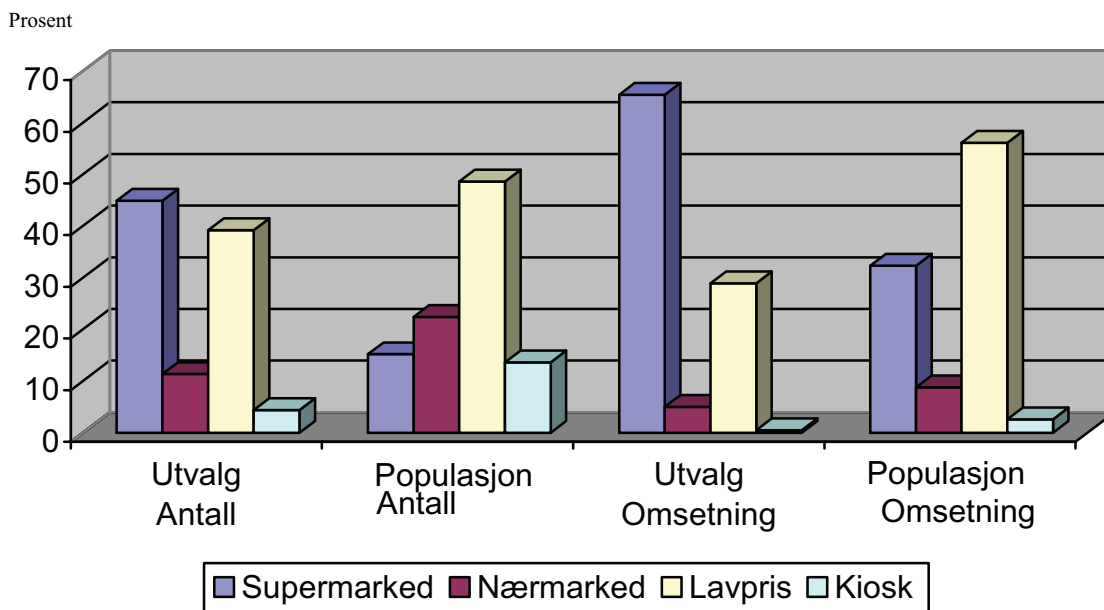
Dagligvarehandelen i Norge er preget av noen få store aktører. Disse aktørene har i lengre tid hatt en stabil markedssituasjon, og deres fokus på konseptutvikling har vært sterkt. Det er gjort analyser basert på det opprinnelige butikkutvalget som viser at kjedenes ulike profiler har en større betydning for prisutviklingen enn butikkens geografiske beliggenhet. Det er derfor fornuftig å stratifisere trekkpopulasjonen inn i konsepter. Vi har valgt å stratifisere etter konseptene kiosk, lavpris, supermarked og nærmarked.

Inndelingen av trekkpopulasjonen i kjedenes konsepter viser at de ulike strata er relativt forskjellig med hensyn til omsetning og antall butikker innen hvert enkelt stratum. Med andre ord har de ulike strata en relativt stor spredning i deres gjennomsnittsomsetning. I allokeringen av den totale utvalgsstørrelsen er det for vårt formål hensiktsmessig å vektlegge at strata som er store og/eller har stor spredning i omsetningen enn små homogene strata. Vi har derfor valgt å benytte den såkalte Neyman allokeringen. Etter at vi har valgt å allokere etter Neyman har vi trukket butikkutvalget *enkelt tilfeldig uten tilbakelegging*.

Etter at butikkutvalget ble trukket ble kjedenes hovedkontor kontaktet for å levere strekkodedata for de utvalgte butikkene til KPI. Deretter ble det nye utvalget korrigert for ulike årsaker for frafall (butikkene ikke har rutiner for å levere data til deres hovedkontor, butikkene er opphørt etc.). Det ble også gjort noen justeringer i utvalget for å også ivareta det geografiske aspektet.

Det vil en gang i året (juli måned) foretas en grundig gjennomgang av det etablerte bedriftsutvalget. Da vil nye butikker trekkes for å inngå i utvalget for å erstatte butikker som har falt fra av ulike grunner i løpet av året. Samtidig må det kontrolleres om hvorvidt nye konsepter er kommet til eller forsvunnet i perioden, og vurdere om eventuelt nye strata skal opprettes.

**Figur 4: Oversikt over konseptfordeling i utvalg og populasjon**



Figur 4 viser supermarkedene utgjør 45 prosent av butikkene i utvalget. Deres omsetningsandel i utvalget er over 60 prosent. Supermarkedskonseptet er overrepresentert sett i forhold til populasjonen. Nærmarkedskonseptet er underrepresentert i utvalget både i antall og omsetning. Ved siden av supermarkeder består utvalget hovedsakelig av lavprisbutikker. De utgjør i underkant av 40 prosent, men har en betydelig mindre omsetningsandel. Lavprisbutikkene er noe underrepresentert i forhold til populasjonen. Nærbutikkene og kioskene er relativt svakt representert i utvalget.

Denne fordelingen skyldes i hovedsak at Neymans allokeringsmetode vektlegger spredningen i omsetningen innenfor hvert enkelt stratum. Vi vet fra tidligere at spredningen i omsetningstallene er stor innenfor supermarkedene. Dermed får butikker innen dette konseptet relativt sett stor sannsynlighet for å bli uttrukket til undersøkelsen. Lavpriskonseptet har betydelig mindre spredning i omsetningen enn supermarkedene, men tildeles relativt mange butikker ved Neymans metode på grunn av deres sterke utbredelse i markedet. Både nærbutikkene og kioskene har relativt sett lav spredning i omsetningen, og får dermed relativt sett lavere sannsynlighet for å bli trukket.

## 6. Oppsummering

Formålet med dette notatet har vært å dokumentere de vurderingene som ble gjort ved etableringen av en ny utvalgsplan for ny delindeksen for matvarer og alkoholfrie drikkevarer. Butikkutvalget trekkes som et sannsynlighetsutvalg, slik at alle butikkene har en kjent sannsynlighet for å bli trukket til butikkutvalget.

Opplegget for utvalgstrekkningen er som følger:

1. BoF er trekkeregisteret for bedrifter i SSB, og benyttes ved trekkingen av butikkutvalget.
2. Det er kun butikker som tilhører kjedene Coop Norge, ICA Norge, NorgesGruppen, Rema1000 og Narvesen denne utvalgsplanen gjelder for.
3. Trekkpopulasjonen stratifiseres etter kjede og konsept.
4. Optimal allokering benyttes i fordelingen av den totale utvalgsstørrelsen.
5. Enkel tilfeldig trekking uten tilbakelegging anvendes som trekkemetode.
6. Årlig gjennomgang av det eksisterende utvalget for å justere for frafall, samt eventuelt etablere nye strata og/eller fjerne etablerte strata.

De ulike kjedene forsøker i stor grad å posisjonere seg i markedet med forskjellige konsept, det vil si med ulik vektlegging av faktorer som pris, beliggenhet, varesortiment og servicenivå. Vi har funnet det hensiktsmessig å stratifisere trekkpopulasjonen inn i konsept. Begrepet konsept er noe diffust ved at de ulike kjedene kan vektlegge ulike ting i sine konsept. Noen viktige forhold i kjedenes konseptutvikling synes å være pris, beliggenhet, varesortiment og servicenivå. Vi ønsker at minst to butikker fra alle kjedene og deres konsepter skal bli trukket til det nye butikkutvalget. Analyser av vårt eksisterende datamaterialet tyder på at lavprisbutikker har en annen prisutvikling enn de andre konseptene. Dette styrker grunnlaget for å stratifisere etter kjede og konsept. En vellykket stratifisering vil øke presisjonen sammenlignet med en trekning uten stratifisering.

Det er fornuftig å fordele den totale utvalgsstørrelsen ut på hvert enkelt stratum ved bruk av en optimal allokering. Til dette har vi valgt å benytte den såkalte Neyman allokeringen. En optimal allokering sikrer at store strata med stor spredning i omsetningen vektlegges i fordelingen av den totale utvalgsstørrelsen.

En allokering basert på Neymans formel forutsetter at butikkutvalget trekkes enkelt tilfeldig uten tilbakelegging innen hvert enkelt stratum.

I forslaget til ny utvalgsplan er det brukt mye ressurser på å kartlegge populasjonen i Bedrifts- og foretaksregisteret, slik at det er mulig å kunne stratifisere etter kjedenes konsept. Et resultat av denne kartleggingen er at det etablert et samarbeid mellom seksjon for bedriftsregister, seksjon for bygg- og tjenestestatistikk og seksjon for økonomiske indikatorer for identifisering av butikkens konsept innen næring 52. Dette identifiseringsarbeidet vil i fremtiden kunne lette kommende utvalgstrekkninger stratifisert etter kjedenes konsept.

Detaljhandelen er en bransje i stadig forandring. Bransjen har i de senere årene fått merke en skjerpet internasjonal konkurranse. Nye internasjonale aktører har etablert seg i Norge. Økt konkurranse har gitt og vil etter all sannsynlighet gi flere endringer i flere av kjedenes butikk konsept. Nye konsepter vil komme til, mens andre konsepter forsvinner. Det vil i fremtiden være nødvendig å følge med nøye i

markedet for detaljhandelen for å fange opp endringer i kjedenes konseptprofiler for å ta hensyn til dette i stratifiseringsarbeidet.

SSBs omsetningsstatistikk over detaljhandelen innhenter omsetningsdata fra de største kjedene innen næring 52.11. Det er tilnærmet fulltelling, men mangler omsetningsdata fra butikker innen en viktig aktør. Dersom dekningsgraden for denne aktøren økes kan data fra detaljomsetningsindeksen benyttes til estimeringene. Ved eventuelle kommende utvalgstrekkinger kan detaljomsetningsindeksen vurderes som et alternativ som trekkeregister for å benytte nyere omsetningstall enn hva som finnes i Bedrifts- og foretaksregisteret.

Vi har valgt å ikke stratifisere etter de åtte geografiske områdene slik det gjøres i øvrige deler av KPI. Beregninger for prisutviklingen etter geografisk område basert på eksisterende butikkutvalg viser ingen signifikante forskjeller i prisutviklingen på matvarer og alkoholfrie drikkevarer fordelt på de ulike geografiske områdene - med unntak av Oslo og Nord- Norge. Derimot kan det være ønskelig å stratifisere på geografiske områder for andre formål.

## Referanseliste

- Bråten, A. (1996): "Populasjon og utvalg - konsumprisindeksen" Notater 96/61, Statistisk sentralbyrå.
- Cochran, W. G. (1977): Sampling Techniques. Wiley & Sons, New York.
- Dagligvarehandelen (2006): "Hvem er hvem i nordisk dagligvarehandel", Utgave 1 - 2006.
- Haraldsen F. og J. Rodriguez (2005): *Den nye matvareindeksen: Bruk av strekkodedata i konsumprisindeksen*, Økonomiske analyser 4/2005, Statistisk sentralbyrå.
- Statistisk sentralbyrå (2001): *Konsumprisindeksen 1995- 2000*., NOS C 680.
- Statistisk sentralbyrå (1994):, *Standard for næringsgruppering*: NOS C 180.
- Särndal, Swensson, Wretman. (1992): Model assisted survey sampling, Springer Verlag, New York.

## De sist utgitte publikasjonene i serien Notater

- 2006/19 A. Holmøy og B.O. Lagerstrøm: Interkommunalt legevaktsamarbeid - en forundersøkelse. 19s.
- 2006/20 H. Tønseth: Årsrapport for kontaktutvalget for helse- og sosialstatistikk. 19s.
- 2006/21 D. Gronna og S. Todsens: Nasjonalregnskap: Beregning av olje- og gassnæringene. 31s.
- 2006/22 D. Gronna, S. Todsens og K. Erlandsen Kolshus: Beregning av olje- og gassnæringene i KNR. 24s.
- 2006/23 T. Tveiekrem Sæter og I. T. Holmen: Prisindeks for bilutleie. 31s.
- 2006/24 Ø. Linnestad og G.M. Molseth: Forprosjekt "Godstransport på kysten". 66s.
- 2006/25 K. Loe Hansen: Indikatorer på kjemikalieområdet- Risiko for skade på helse og miljø grunnet bruk av kjemiske stoffer. 46s.
- 2006/26 A. Akselsen, S. Lien og Ø. Sivertstøl: FD-Trygd. Variabelliste. 58s.
- 2006/27 J. Heldal og A. Rusti: Om samordning av utvalg vedbruk av PRN-tall. 29s.
- 2006/28 C. Nordseth og Ø. Sivertstøl: FD - Trygd. Dokumentasjonsrapport. Fødsels- og sykepenger, 1992-2003. 134s.
- 2006/29 A. Linderud: Verdipapirstatistikk. Dokumentasjonsnotat. 54s.
- 2006/30 V.V. Holst Bloch, H. Høye, M. Steinnes og J.K. Undelstveit: Kartbasert rapportering i KOSTRA - en mulighetsstudie. 50s.
- 2006/31 E. Høydal: Monitor for sekundærflytting. En deskriptiv analyse av sekundærflyttinger blant flyktninger bosatt i Norge i 1995-2004. 67s.
- 2006/32 E.Cometa Rauan: Undersøking om foreldrebetaling i barnehagar, januar 2006. 46s.
- 2006/33 T. Skarøhamar: Kriminalitet gjennom ungdomstiden blant nordmenn og ikke-vestilige innvandrere. En analyse av fødselskullet 1977. 36s.
- 2006/34 N. Hagesæther og L-C. Zhang: Om arbeidsledighet i AKU og Arena. 19s.
- 2006/35 T. Hægeland, Lars J. Kirkebøen og Oddbjørn Raaum: Skoleresultater 2005. En kartlegging av karakterer fra grunnskoler og videregående skoler i Norge. 83s.
- 2006/36 S. Skaare: Undersøkelse om «Utbrenthet i enkelte yrker» 2005. Dokumentasjonsrapport. 68s.
- 2006/37 O.F. Vaage: Barn og unges idrettsdeltakelse og foreldres inntekt. Analyse med data fra Levekårsundersøkelsen 2004. 31s.
- 2006/38 A.Vedø og L. Solheim: En praktisk innføring i utvalgsplanlegging. 40s.
- 2006/39 H.C. Hougen: Samordnet levekårsundersøkelse 2005 - tverrsnittundersøkelsen. Dokumentasjonsrapport. 156s.
- 2006/40 T. Nøtnæs, S. Bytingsvik og B. Hole: Resultater fra brukertesting av ssb.no. 34s.
- 2006/41 KOSTRA. Arbeidsgrupperapporter 2006. 169s.
- 2006/42 T. Gulbrandsen: Levekårsundersøkelse blant studenter. Dokumentasjonsrapport. 66s.
- 2006/43 A-G. Jørstad: Overvåkingssystemet for bedrifter i Bof. 19s.
- 2006/44 M. høstmark og B.O. Lagerstrøm: Undersøkelse om Arbeidsmiljø: Destruktiv atferd i arbeidslivet. Dokumentasjonsrapport. 43s.
- 2006/45 T.K. Schjerven og K.Å. Wass: Faglig modell og rammeverk i StatRes. 67s.