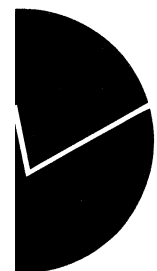


Arnfinn Schjalm

**Sluttrapport om utvalg og
estimering for kulturland-
skapsovervåking**

Notater



1. Sammendrag

Statistisk sentralbyrå har på oppdrag fra Norsk institutt for jord- og skogkartlegging gitt anbefalinger om utvalgsplan og estimeringsopplegg for flyfotobasert overvåking av jordbrukets kulturlandskap.

SSB anbefaler at Landskogstakseringens utvalg av ruter med jordbruksland i senterpunktet benyttes.

Det anbefales at alle punkt-, linje- og arealelementer estimeres med en regresjonsestimator med registrert jordbruksland som hjelpevariabel. Det svarer til å bruke vektene $w_i^{(3)}$ beskrevet i kapittel 5. For estimering av indeksvariable anbefales direkte innsetting av estimater for punkt-, linje- og arealestimater.

For estimering av variansen til punkt-, linje- og arealestimatene anbefales en Jackknifeestimator, mens variansen for indeksvariable avledes av disse ved hjelp av en formel basert på Taylorutvikling.

2. Bakgrunn og saksgang

Norsk institutt for jord- og skogkartlegging (NIJOS) har ansvaret for "System for tilstandsovervåking og resultatkontroll av jordbrukets kulturlandskap", basert på tolkning av flyfotografier av et utvalg av 1-km²-ruter. Statistisk sentralbyrå (SSB) har bistått med valg av utvalgsplan og estimeringsopplegg. I denne sluttrapporten gis en kort oppsummering av saksgangen og de konklusjoner SSB har gitt i forbindelse med utvalg og estimering. Saksgangen beskrives i dette kapitlet, den utvalgsplanen som ble valgt i kapittel 3, problemstillinger som ble identifisert i forbindelse med estimeringen i kapittel 4, og anbefalinger om estimeringsopplegget og noen vurderinger knyttet til dem i kapittel 5. I kapittel 6 gis noen anbefalinger i forbindelse med behandlingen av data fra feltkontroll. SSBs notat av 26. mars 1998 er tatt med som vedlegg.

Statistisk sentralbyrå påtok seg i april 1997 som eksternt oppdrag å gi faglige kommentarer til et notat (Strand, 1997) fra NIJOS der utvalgsplan og estimatorer for kulturlandskapsovervåking ble drøftet. Som svar på dette ble det i mai 1997 utarbeidet et notat (Schjalm, 1997) med kommentarer. I 1998 fikk vi et nytt eksternt oppdrag fra NIJOS om utarbeiding av utvalgsplan og estimering for kulturlandskapsovervåkingen. Prosjektet er beskrevet i (NIJOS, 1998). Det ble holdt to møter mellom NIJOS og SSB i februar 1998. Den viktigste problemstillingen på det tidspunktet var valg av utvalgsplan, som måtte være klar før NIJOS kunne gjøre ferdig anbudsspesifikasjon for flyfotograferingen. Valget sto mellom å bruke et utvalg av ruter med jordbruksland i senterpunktet som Landsskogstakseringen hadde identifisert eller å trekke et nytt utvalg der registerinformasjon om landbruksbygninger ble utnyttet. Fordeler og ulemper ved disse utvalgsplanene ble vurdert i (Schjalm, 1998a) og bruk av Landskogstakseringens utvalg anbefalt. Denne utvalgsplanen ble valgt av NIJOS, og flyfotograferingen ble påbegynt sommeren 1998. På et møte mellom NIJOS og SSB i august 1998 ble det avtalt hvilke punkter som skulle fokuseres i et notat om estimeringen. Dette notatet ble levert i desember 1998 (Schjalm, 1998b).

3. Utvalgsplan

3.1 Beskrivelse av utvalget og utvalgsplanen

Det ble bestemt å bruke det utvalget på omlag 1 400 ruter med jordbruksland i senterpunktet som Landsskogstakseringen har identifisert. Dette utvalget kan betraktes som en realisering av følgende tottrinns utvalgsplan: På første trinn trekkes i hver av 8 soner et 3x3-km forband av ruter på 1x1 km. Rutene i forbandet ligger systematisk, slik at hver tredje rute i øst-vest-retningen og hver tredje rute i nord-syd-retningen er med i forbandet. Forbandene dekker dermed omlag 1/9 av arealet i hver sone.

Det uttrukne forbandet i hver sone vil vi betrakte som et av 9 mulige forband trukket med sannsynlighet 1/9 og forbandene kan betraktes som at de er trukket uavhengig. Av rutene i disse

forbandene er de rutene tatt med i utvalget som har jordbruksland i senterpunktet. Dette betrakter vi som at det er gjort en trekking på annet trinn med trekkesannsynlighet lik antall kvadratkilometer jordbruksland i ruten. Trekkesannsynligheten for par av ruter er ikke like opplagt, men det virker rimelig å bruke en antagelse om uavhengig trekking når variaser skal beregnes. For denne utvalgsplanen foreligger det allerede et utvalg. Landsskogtakseringen har identifisert et utvalg på 1 381 ruter utenom Finnmark. For å fullføre identifiseringen av dette utvalget må senterpunktene i rutene i Finnmark undersøkes. Det ventes å gi omlag 15 ruter til.

Eventuelle variasjoner av utvalgsplanen vil bestå i å supplere dette utvalget. Mest aktuelt er i så fall ekstra ruter i Finnmark, Aust-Agder og fjellområdene, for å få bedre estimater på regionnivå.

3.2 Egenskaper ved utvalgsplanen

3.2.1 Mulig systematikk i landskapet

I (Schjalm, 1997) ble NIJOS gjort oppmerksom på den generelle risikoen for systematikk i populasjonen som er knyttet til systematisk utvalg. Spesielt vil systematiske underutvalg av 3x3-km forbandene ikke være ønskelig, men det er det heller ikke noe behov for. NIJOS opplyser at det er en avstandsavhengig systematikk i norsk natur med periode omkring 30 km. NIJOS regner det som mindre trolig at det skulle finnes en systematikk med periode på 3 km, men vil ikke helt utelukke det. Gårdene i mange norske daler ligger litt oppe i lia på hver side av dalen, og en kan ikke utelukke at avstanden kan være omtrent 3 km i noen tilfeller.

Det antas at 3x3-km forband til tross for dette er akseptabelt. Selv om det skulle være systematikk i utvalget innenfor hvert dalføre, vil utvalget neppe treffe likt i alle dalfører. Derfor er det grunn til å regne med at de ulike typer områder som kan finnes i et dalføre vil bli dekket av utvalget.

3.2.2 Manglende estimator for førstetrinnsvarians

Et annet problem med det systematiske utvalget er at det ikke tillater forventningsrett estimering av utvalgsvariansen på første trinn.

3.2.3 Moderat korrelasjon mellom rutene i utvalget

Det systematiske utvalget inneholder ikke naboruter. Da naboruter må antas å være høyere korrelert enn andre ruter, er det sterk grunn til å tro at rutene i et systematisk utvalg er lavere korrelert enn et tilsvarende enkelt tilfeldig utvalg når det gjelder areal av forskjellig type og forekomst av forskjellige fenomener. Med «tilsvarende» tenkes her på kravet om jordbruksland i senterpunktet. Lavere korrelasjon mellom rutene innebærer at utvalget gir mer informasjon.

3.2.4 Dekningen av jordbruksland

Siden sannsynligheten for jordbruksland i senterpunktet øker med mengden jordbruksland i ruten, vil annet trinn i utvalgsplanen resultere i større gjennomsnittlig jordbruksareal per rute i utvalget enn et enkelt tilfeldig utvalg, og derfor gi større datagrunnlag for å estimere fordelingen av jordbruksland enn et enkelt tilfeldig utvalg av samme størrelse. Forskjellen i forventet jordbruksareal per rute er bestemt av fordelingen av jordbruksareal i ruter med jordbruksland.

3.2.5 Fenomener i områder med tynt jordbruksland

På den annen side vil denne utvalgsplanen gi liten dekning av fenomener som spesielt er knyttet til områder med et lavt innslag av jordbruksland.

3.2.6 Ikke behov for å etablere utvalget

En praktisk fordel med den utvalgsplanen som ble valgt var at det ikke var nødvendig å gjøre noe arbeid for å etablere utvalget, idet Landsskogtakseringen hadde identifisert ruter med jordbruksland i senterpunktet. Dette gjelder ruter som hadde jordbruksland i senterpunktet da Økonomisk Kartverk

ble etablert eller ajourført. Tidspunktet for dette varierer, men kan ligge helt tilbake til 1960-tallet. Situasjonen kan derfor være litt annerledes nå, men det gjøres neppe noen vesentlig feil ved å overse dette avviket.

3.2.7 Utvalgsstørrelse

Landsskogtakseringen har identifisert 1 381 ruter utenom Finnmark med jordbruksland i senterpunktet. En tilsvarende gjennomgang for Finnmark ventes å gi omlag 15 ruter til.

Det er mulig å supplere utvalget med ekstra ruter i regioner hvor det ønskes mer presise estimater enn Landskogstakseringens utvalg tillater. Finnmark, Aust-Agder og fjellområdene har vært nevnt som aktuelle områder for tilleggsutvalg.

4. Problemstillinger knyttet til estimering

Det ble avtalt å fokusere følgende problemstillinger knyttet til estimering:

4.1 Estimering av landskapsvariablene

Det skal utarbeides et estimeringsopplegg for fire typer variable, nemlig punktelementer, linjeelementer, arealelementer og indekser. Det tas sikte på å lage et felles estimeringsopplegg for elementer av samme type. Indekser er funksjoner av de andre typene. Især for indeksene, men også for de andre typene variable kan det tenkes forhold som tilsier en ytterligere differensiering av estimeringsopplegget. NIJOS leverte en liste over de aktuelle variablene (se NIJOS, 1998, s. 25-26). Registerinformasjon, især informasjon om jordbruksareal er aktuelle hjelpevariabler ved estimeringen.

4.2 Varians

Det ønskes usikkerhetsmål knyttet til estimatene. Et spesielt problem blir førstetrinnsvariansen i utvalgsplanen.

4.3 Elementer utenom jordbruksland

Utvalget består av ruter *med jordbruksland i senterpunktet*. Det fører til at sannsynligheten for at et punkt P er med i utvalget er proporsjonal med arealet av jordbruksland i en 1-km²-rute med senterpunkt i P. Det er således grovt sett en avtakende inklusjonssannsynlighet med økende avstand fra jordbruksland. Dette må det tas hensyn til ved estimeringen, i alle fall for estimering av elementer utenfor jordbruksland. Informasjon om avstand til jordbruksland i ruten vil være tilgjengelig, og kan være aktuell å bruke ved estimeringen.

4.4 Målefeil

For kontrollrutene vil det for et grid på 100 punkter per rute med avstand 100 meter foreligge opplysninger om arealtype i punktet både fra flyfototolkningen og feltarbeidet. Dette gir opplysninger om målefeil ved flyfototolkningen som bør utnyttes ved estimering og variansberegning. Det utgjør også et grunnlag for kunnskap om slike målefeil.

4.5 Ruter nær sonегrensene

Utvalget av ruter er uavhengige utvalg av ruter i 8 soner. Nær sonегrensene kan rutene ha større eller mindre avstand enn vanlig. Det må avklares om dette har nevneverdig betydning for estimeringen, om det bør brukes spesielle vektorer for ruter nær sonегrensene eller andre justeringer i estimeringen. NIJOS kan bidra med opplysninger om samlet lengde av sonегrenser o.l.

5. Estimering

5.1 Generell estimeringsstrategi

Totaltall for arealvariable, linjevariable og punktvariable estimeres ved å blåse opp observerte størrelser. Vektene i oppblåsing tar utgangspunkt i inklusjonssannsynlighetene, men forskjellige justeringer kan være aktuelle. Uten justeringer og med riktige inklusjonssannsynligheter får vi Horwitz-Thompson estimatoren. Indeksvariablene, som er forskjellige funksjoner av areal-, linje- og punktvariable estimeres deretter ved å sette de estimerte verdiene for areal-, linje- og punktvariablene inn i disse funksjonene. Fremgangsmåten blir den samme for regioner og eventuelle andre delpopulasjoner som for totaltall for hele landet, og gir konsistente resultater med mulig unntak for indeksvariablene.

Variansberegninger for areal-, linje- og punktvariable gjøres med en jackkniveestimator innenfor hver sone, og disse variansestimaterne summeres over sonene. Økningen i variansen nær sonegrensene er så liten at vi ikke trenger å ta hensyn til den i beregningene. Deretter kan variansestimater for indeksvariablene beregnes ved Taylorlinearisering. (Det er heller ikke aktuelt å bruke opplysninger om sone i selve estimeringen.)

Det gjenstår å bestemme vektene. Tre problemstillinger som melder seg når vektene skal fastsettes er disse:

1. Inklusjonssannsynlighetene varierer for punkter i samme utvalgsrute og kan ikke bestemmes nøyaktig.
2. Små inklusjonssannsynligheter gir store oppblåsningsfaktorer i Horwitz-Thompson estimatoren med økt varians som en konsekvens.
3. Det finnes registerinformasjon om dyrkingsareal som bør utnyttes ved estimeringen.

Nedenfor følger en mer detaljert beskrivelse av dette estimeringsopplegget.

5.2 Estimering av arealer, linjeelementer og punktelementer

Vi innfører følgende notasjon:

T - en type landskapselement, f. eks. fulldyrka åker, blandingsskog, steingjerde eller frittstående tre

s_T - samlingen av enheter med landskapselement T

$y_{T,i}$ - variabel, enten antall, lengde eller areal

I utvalget finnes et sett s_T med enheter av hver type landskapselement T det ønskes et estimat for. For hver enhet i i utvalget s_T hører et tall $y_{T,i}$. Mengdene s_T og størrelsene $y_{T,i}$ kan være definert på forskjellige måter for forskjellige landskapselementer. En mulighet er at s_T er mengden av utvalgsruter som inneholder landskapselementet T og at $y_{T,i}$ er antall ganger dette landskapselementet forekommer i rute i . Alternativt kan $y_{T,i}$ være samlet areal eller lengde av forekomstene av landskapselement T i rute i . En annen mulighet er at enhetene i s_T er mindre enn en utvalgsrute, f. eks. en teig eller den delen av en teig som ligger i en utvalgsrute. Da kan $y_{T,i}$ være arealet av teigen (som faller innenfor utvalgsruten) eller et størrelsesmål for landskapselementet T i teigen eller den delen av teigen som ligger innenfor utvalgsruten. I alle tilfeller er den størrelsen vi ønsker å estimere

$$(1) Y_T = \sum_{i \in U_T} y_{T,i},$$

der U_T er mengden av enheter i populasjonen som inneholder landskapselementet T . Vi bruker en estimator av formen

$$(2) \hat{Y}_T = \sum_{i \in s_T} w_{T,i} y_{T,i}.$$

Det gjenstår å fastsette om enhetene i s_T skal være utvalgsruter eller mindre enheter og å bestemme vektene $w_{T,i}$. Medmindre det er ønskelig å organisere dataene slik at det gir en beskrivelse av enkeltteiger eller liknende, er et eventuelt ønske om å variere vektene $w_{T,i}$ innenfor en utvalgsrute eneste grunn til å bruke mindre enheter enn utvalgsrutene.

Dataene bør organiseres med én record for hver enhet i , og med w_T og y_T som variable. Opplysninger om y_T for flere landskapselementer T kan naturligvis ligge på samme record og w_T kan være felles for flere landskapselementer T , muligens alle. Dessuten må den inneholde variabler for sone og rute og også ident for i hvis enheten ikke er rute.

5.3 Estimering av indeksvariable

Indeksvariablene har formen $I = k \cdot Y^\alpha \cdot Z^\beta$, der Y og Z er variable som estimeres som beskrevet i avsnitt 2.2, og k , α og β er konstanter. De estimeres ved innsetting av estimatene for Y og Z , altså $\hat{I} = k \cdot \hat{Y}^\alpha \cdot \hat{Z}^\beta$.

Noen eksempler på indeksvariable er andel av jordbruksareal som er av en bestemt type (forhold mellom to arealer), andel av gjerdelengde som er av en bestemt type (forhold mellom to lengdemål) og tetthet av stier (dvs. antall km stier per km² landskap; forhold mellom et lengdemål og et areal).

Spesielt nevnes “shape index” for en teig, som er definert som $I_{shape} = (4\pi)Y / Z^2$, der Y er arealet og Z omkretsen av teigen. Shape index måler hvor uregelmessig form teigen har, tar en verdi mellom 0 og 1, og antar maksimal verdi når teigen har sirkelform.

5.4 Vekting og kalibrering

Vektene w_i i estimeringen kan velges på flere måter. Nedenfor beskrives tre muligheter. Vi lar U_k betegne alle rutene innenfor et kalibreringsområde k , f. eks. fylke, og s_k utvalgsrutene i det samme området. N_k er antall ruter i U_k , og n_k antall ruter i s_k . Vi lar J_i betegne observert jordbruksareal i rute i , og X_k betegne registrert jordbruksareal i område k . For de enkelte ruter må observert og registrert jordbruksareal gjelde samme år, nemlig det året ruten er fotografert, men det er ikke strengt nødvendig at alle utvalgsruter i et område er fotografert samme år. Vi kaller k et kalibreringsområde fordi vektene i to av alternativene nedenfor velges slik at jordbruksareal i hvert område k blir riktig estimert.

5.4.1 Horvitz-Thompson estimatoren

En enkel og alminnelig måte å lage vektorer på er å bruke inverse trekkesannsynligheter. Det gir Horvitz-Thompson estimatoren. Betrakter vi en ruteinndeling av landet som gitt og antar at sannsynligheten for jordbruksland i senterpunktet av en rute er det samme som andelen jordbruksland i ruten, får vi at trekkesannsynligheten for rute i er $\frac{J_i}{9}$, og den inverse trekkesannsynligheten blir

$$(3) w_i^{(1)} = \frac{9}{J_i}.$$

Det er flere ulemper med å bruke disse vektene her. Det er mulighet for store variasjoner i vektene på grunn av forskjellene i arealet med jordbruksland i rutene. Stor variasjon i vektene gir stor varians i de estimerte størrelsene. Det er faktisk mulig at $J_i=0$, siden jordbruksvirksomheten i ruten kan ha opphørt siden Landskogstakseringens fotografering. En mulig måte å håndtere disse problemene på er å bruke

$$(4) \quad w_i^{(1,M)} = \min\left(\frac{9}{J_i}, M\right)$$

istedenfor $w_i^{(1)}$, der M er en valgt maksimumsverdi.

Den viktigste grunnen til ikke å bruke inverse trekkesannsynligheter for estimering av variable med tilknytning til jordbruksland er at de ikke utnytter registerinformasjonen om jordbruksareal. Nedenfor følger to forslag til vektorer som utnytter denne registerinformasjonen.

5.4.2 Kalibrerte vektorer

Ved å multiplisere vektene $w_i^{(1)}$ i et område k med en felles konstant φ_k bestemt ved

$$\varphi_k \sum_{i \in S_k} w_i^{(1)} J_i = X_k, \text{ får vi vektene}$$

$$(5) \quad w_i^{(2)} = \varphi_k w_i^{(1)} = \frac{X_k}{9n_k} \cdot \frac{9}{J_i} = \frac{X_k}{n_k J_i}.$$

Bruk av disse vektene vil gi et korrekt estimat for jordbruksarealet og derfor også et godt estimat for variable som er sterkt korrelert til variabelen jordbruksareal. Men fortsatt kan vektene variere mye med tilhørende varians i estimatet.

Et alternativ er å bruke like vektorer for alle utvalgsruter i et område, men kalibrert slik at jordbruksareal estimeres riktig. Vektene blir da

$$(6) \quad w_i^{(3)} = \frac{X_k}{\sum_{i \in S_k} J_i}.$$

Disse vektene svarer til følgende modell:

$$y_i = \beta J_i + \varepsilon_i, \text{ der } E(\varepsilon_i) = 0 \text{ og } Var(\varepsilon_i) \propto J_i.$$

Estimering av β i denne modellen med veiede minste kvadraters metode gir

$$\hat{\beta} = \frac{\sum_{i \in S_k} y_i}{\sum_{i \in S_k} J_i}.$$

$$\text{Det gir } \hat{Y} = \sum_{i \in U_k} \hat{y}_i = \sum_{i \in U_k} \frac{\sum_{j \in s_k} y_j}{\sum_{j \in s_k} J_j} x_i = \frac{\sum_{i \in s_k} y_i}{\sum_{i \in s_k} J_i} \cdot \sum_{i \in U_k} x_i = \frac{X_k}{\sum_{i \in s_k} J_i} \cdot \sum_{i \in s_k} y_i = w_i^{(3)} \sum_{i \in s_k} y_i .$$

Anbefaling:

Dette siste alternativet, altså $w_i^{(3)}$, er det vi vil anbefale, i alle fall for variable korrelert med jordbruksareal. Vektene er kalibrert med hensyn på jordbruksland, men varierer likevel ikke for mye. Vektene $w_i^{(2)}$ tar i større grad hensyn til at inklusjonssannsynlighetene for de enkelte punkter varierer i takt med mengden jordbruksland i nærområdet. Fenomener med svakere tilknytning til jordbruksland kan bli noe bedre estimert med $w_i^{(1)}$ eller $w_i^{(2)}$. Men fordi disse kan ha meget stor varians, anbefaler vi likevel $w_i^{(3)}$ for alle variabler.

5.4.3 Kalibreringsområder

Ved bruk av vektene $w_i^{(2)}$ eller $w_i^{(3)}$ må det velges en inndeling av populasjonsrutene i kalibreringsområder. Brukes kommuner blir det bare 4 til 5 utvalgsruter i gjennomsnitt per kalibreringsområde, noe som kan gi stor variasjon i vektene. En inndeling i fylker synes mer passende. En inndeling i et liknende antall homogene jordbruksregioner kan være et alternativ til å dele inn etter fylke.

Anbefaling:

Vi anbefaler å bruke fylkene som kalibreringsområder.

5.5 Estimering av varianser

Utvalgsvarianser for estimatene i avsnitt 2.2 estimeres med Jackknifeestimatoren. Av de variantene som finnes for denne estimatoren velges den som angir høyest varians, fordi det i den foreliggende estimeringssituasjonen foreligger utvalgsvariasjon som ikke blir tatt hensyn til, nemlig plasseringen av forbandet av senterpunkter i hver sone.

Jackknifeestimatet beregnes slik:

For hver utvalgsrute $i \in s$ estimeres parameteren Y_T basert på et dataset med rute i fjernet, altså basert på utvalget $s \setminus \{i\}$, men forøvrig på samme måte som ved beregningen av \hat{Y}_T . Vi betegner resultatet av denne beregningen $\hat{Y}_{T,-i}$. Jackknifeestimatet for variansen til \hat{Y}_T er da

$$(7) \hat{V}_{JK}(\hat{Y}_T) = \frac{1}{n(n-1)} \sum_{i \in s} (\hat{Y}_{T,-i} - \hat{Y}_T)^2, \text{ der } n \text{ er antall ruter i utvalget } s.$$

Utvalgsvarianser for indeksestimater av formen $\hat{I} = k \cdot \hat{Y}^\alpha \cdot \hat{Z}^\beta$ beregnes ved hjelp av Taylorutvikling og baseres på variansestimater for \hat{Y} , \hat{Z} og $\hat{Y} + \hat{Z}$, beregnet ved ovenstående metode. Det gir følgende formel for variansen til \hat{I} :

$$(8) \hat{V}_{Taylor}(\hat{I}) = \hat{I}^2 \left[\left(\frac{\alpha^2}{\hat{Y}^2} - \frac{\alpha\beta}{\hat{Y}\hat{Z}} \right) \hat{V}_{JK}(\hat{Y}) + \left(\frac{\beta^2}{\hat{Z}^2} - \frac{\alpha\beta}{\hat{Y}\hat{Z}} \right) \hat{V}_{JK}(\hat{Z}) + \frac{\alpha\beta}{\hat{Y}\hat{Z}} \hat{V}_{JK}(\hat{Y} + \hat{Z}) \right].$$

Innsetting av $\alpha = 1$ og $\beta = -1$ gir følgende variansestimater for en rate eller en andel:

$$(9) \hat{V}_{Taylor}(\hat{Y} / \hat{Z}) = \hat{I}^2 \left[\left(\frac{1}{\hat{Y}^2} + \frac{1}{\hat{Y}\hat{Z}} \right) \hat{V}_{JK}(\hat{Y}) + \hat{V}_{JK}(\hat{Y}) - \frac{1}{\hat{Y}\hat{Z}} \hat{V}_{JK}(\hat{Y} + \hat{Z}) \right].$$

Innsetting av $\alpha = 1$ og $\beta = -2$ gir varians til estimatet for shape index:

$$(10) \hat{V}_{Taylor}(k\hat{Y} / \hat{Z}^2) = \hat{I}^2 \left[\left(\frac{1}{\hat{Y}^2} + \frac{2}{\hat{Y}\hat{Z}} \right) \hat{V}_{JK}(\hat{Y}) + \left(\frac{4}{\hat{Z}^2} + \frac{2}{\hat{Y}\hat{Z}} \right) \hat{V}_{JK}(\hat{Z}) - \frac{2}{\hat{Y}\hat{Z}} \hat{V}_{JK}(\hat{Y} + \hat{Z}) \right].$$

5.6 Ruter nær soneregrensene

Når et senterpunkt ligger mindre enn 500 meter fra en soneregrense vil ruten med dette senterpunktet strekke seg inn i en nabosone, og kan overlape en utvalgsrute med senterpunkt i den sonen. Det er valgt å bruke hele utvalgsruten også når den gjennomskjæres av en soneregrense. Utvalgsplanen tillater derfor at områder nær soneregrensene kommer med i to utvalgsruter, noe som ikke skjedde for det utvalget som faktisk ble trukket. Følgende beregninger viser imidlertid at det ikke er nødvendig å ta hensyn til den mulige overlappingen av utvalgsruter nær soneregrensene, hverken ved estimeringen eller ved beregningen av utvalgsvarianser.

La P være et punkt som ligger mindre enn 500 m fra en soneregrense og J_1 og J_2 være arealet av jordbruksland målt i km^2 i de to delene av en 1-km^2 -rute med senterpunkt i P (og riktig orientering) som ligger på hver sin side av en soneregrense. P kommer med i utvalget hvis et punkt med jordbruksland i denne 1-km^2 -ruten blir trukket ut som et senterpunkt. P kommer med to ganger hvis to punkter med jordbruksland i denne ruten trekkes ut som senterpunkt, altså ett fra hver sone. La I betegne antall senterpunkter som trekkes fra denne ruten.

Vi har da at $J_1 + J_2 \leq 1$ og følgende sannsynlighetsfordeling for I:

$$P(I = 0) = \left(1 - \frac{J_1}{9} \right) \left(1 - \frac{J_2}{9} \right),$$

$$P(I = 1) = \frac{J_1}{9} \left(1 - \frac{J_2}{9} \right) + \left(1 - \frac{J_1}{9} \right) \cdot \frac{J_2}{9},$$

$$P(I = 2) = \frac{J_1}{9} \cdot \frac{J_2}{9}.$$

Vi kan nå beregne forventningen til I:

$$E(I) = \frac{J_1}{9} \left(1 - \frac{J_2}{9} \right) + \left(1 - \frac{J_2}{9} \right) \cdot \frac{J_2}{9} + 2 \cdot \frac{J_1}{9} \cdot \frac{J_2}{9} = \frac{J_1 + J_2}{9} = \frac{J}{9}.$$

Vi ser at selv om inklusjonssannsynligheten til punktet P er litt mindre enn $\frac{J}{9}$, er forventet antall

ganger P inkluderes eksakt $\frac{J}{9}$. Bruker vi Horvitz-Thompson-vektorer uten å ta hensyn til soneregrensene ved beregning av inklusjonssannsynlighetene blir bidraget fra ruten i

$$\frac{y_i}{\hat{\pi}_i} \cdot I = \frac{y_i}{\left(\frac{J}{9}\right)} \cdot I .$$

Forventet bidrag fra rute i blir da $E\left(\frac{y_i}{\left(\frac{J}{9}\right)} \cdot I\right) = \frac{y_i}{\left(\frac{J}{9}\right)} \cdot \frac{J}{9} = y_i$, og følgelig blir

$E\left(\sum_{i \in U} w_i y_i\right) = \sum_{i \in U} y_i$. Horvitz-Thompson estimatoren forblir altså forventningsrett når vi overser sonegrensene ved beregningen av inklusjonssannsynlighetene.

Følgende beregning gir variansen.

$$E(I^2) = E(I) + (2^2 - 2) \cdot \frac{J_1}{9} \cdot \frac{J_2}{9} = \frac{J}{9} + \frac{2J_1J_2}{81} .$$

$$\text{Da blir } \text{Var}(I) = \frac{J}{9} - \frac{J^2}{81} + \frac{2J_1J_2}{81} .$$

Sonegrensen gir en ekstra varians på $\frac{2J_1J_2}{81}$.

Siden $J_1J_2 = \frac{J_1}{J} \cdot \frac{J_2}{J} \cdot J^2 \leq \frac{1}{4} \cdot J^2$, får vi følgende ulikhet for forholdet mellom denne ekstravariansen og variansen uten sonegrense:

$$\frac{\frac{2J_1 \cdot J_2}{81}}{\frac{J}{9} - \frac{J^2}{81}} \leq \frac{\frac{162}{81}}{\frac{J}{9} - \frac{J^2}{81}} = \frac{J^2}{18J - 2J^2} = \frac{J}{18 - 2J} \leq \frac{J}{16} \leq \frac{1}{16} , \text{ siden } J \in [0,1] .$$

Innenfor en stripe på 1 km bredde langs sonegrensene er det altså en økning av variansen på mindre enn 1/16 av den ordinære variansen. Økningen blir enda mindre mot utkanten av stripene, og når andelen jordbruksland i området er mindre enn 100 %. Alt i alt er det klart at økningen av varians på grunn av sonegrensene kan neglisjeres.

6. Målefeil

Her presenteres en hovedidé for presentasjonen av kontrollundersøkelsen, og noen tanker om momenter som krever videre utvikling.

Et underutvalg av ruter er ved enkel tilfeldig trekking valgt ut for feltkontroll i samme sesong som de fotograferes. 100 punkter i hver av disse rutene blir klassifisert både ved flyfototolkning og ved feltundersøkelse. Klassifikasjonen ved feltundersøkelsen må regnes som fasitt. En naturlig presentasjonsmåte for resultatene av denne undersøkelsen er å lage en krysstabell med flyfotoklassifikasjonen i forspalten og feltklassifikasjon i tabellhodet. I hver celle i tabellen angis antall punkter $A_{m,n}$ med flyfotoklassifikasjon m og feltklassifikasjon n . Diagonalen, som inneholder de riktig klassifiserte punktene kan fremheves med fete typer.

Fra denne tabellen bør det lages en ny tabell ved å prosentuerer hver linje i den første tabellen. Et estimat for andelen av punktene som er plassert i klasse m ved flyfototolkningen som virkelig hører til

klasse n er $a_{m,n} = \frac{A_{m,n}}{\sum_v A_{m,v}} = \frac{A_{m,n}}{A_m}$. Prosentandelen er $100a_{m,n} \%$.

En tenkelig anvendelse av den siste tabellen er å bruke den til omregning av resultater fra flyfototolkning til resultater som tar hensyn til feiltolkninger ved lesing av bildene. En annen mulig bruk er å bruke den til å beregne varians som skyldes tolkningsusikkerhet.

De ideene som er presentert her er svært enkle. De tar ikke hensyn til kunnskap om at visse klassifikasjoner ikke er aktuelle i visse områder. En annen ting er at flyfototolkerne kan ha individuelle feilmønstre. I prinsippet gjelder dette også for dem som gjør feltstudiene, men her er det grunn til å tro at de individuelle forskjellene er små. Datafilen bør inneholde identifikasjon av flyfototolkeren og felttolkeren for hvert punkt.

Her er tenkt at samme punkt klassifiseres med to forskjellige observasjonsteknikker, der den ene, feltstudium, er mye sikrere enn den andre. En mulig kilde til forskjellig klassifikasjon er imidlertid at de to punktene som sammenliknes ikke er sammenfallende på grunn av en eller annen form for "navigasjonsfeil".

Referanser:

NIJOS (1998): *System for "Tilstandsovervåking og resultatkontroll av jordbrukets kulturlandskap"*, femte utkast. Norsk institutt for jord- og skogkartlegging, 2. februar 1998.

Schjalm, Arntfinn (1997): *Notat om Kulturlandskapsovervåking*. Statistisk sentralbyrå, 16. mai 1997.

Schjalm, Arntfinn (1998a): *Notat om utvalgsplan for kulturlandskapsovervåking*. Statistisk sentralbyrå, 26. mars 1998.

Schjalm, Arntfinn (1998b): *Notat om estimering for kulturlandskapsovervåking*. Statistisk sentralbyrå, 14. desember 1998.

Strand, Geir-Harald (1997): *Utvalgs- og Beregningsmetoder for Kulturlandskapsovervåking*. Norsk institutt for jord- og skogkartlegging, 23. april 1997.

VEDLEGG

Sch 26.03.98

Notat om utvalgsplan for kulturlandskapsovervåking

1. Bakgrunn

Norsk institutt for jord- og skogkartlegging (NIJOS) har fått ansvaret for «System for tilstandsovervåking og resultatkontroll av jordbrukets kulturlandskap», basert på tolkning av flyfotografier av et utvalg av 1-km² ruter. NIJOS har gitt Statistisk sentralbyrå (SSB) i oppdrag å utarbeide utvalgsplan og estimeringsopplegg. Vårt arbeid bygger på en tidligere vurdering av alternative utvalgsplaner, (Schjalm, mai 1997). Videre har vi bygget vår utredning på et notat fra NIJOS, (Strand, april 1997). Det er også avholdt to møter mellom SSB og NIJOS.

Formålet med dette notatet er å beskrive og vurdere de utvalgsplaner som nå er aktuelle for flyfotograferingen. Beskrivelsen gis i avsnitt 2, vurderingen i avsnitt 3. Estimeringen omtales i avsnitt 4, og i avsnitt 5 drøftes utvalgsstørrelsen. I avsnitt 6 konkluderer vi med anbefalinger om utvalgsplan og estimeringsopplegg.

2. Beskrivelse av de utvalgsplaner som er til vurdering

To hovedalternativer av utvalgsplaner er til vurdering, men det er visse variasjonsmuligheter innenfor hvert alternativ. Alternativ A er å bruke det utvalget på omlag 1 400 ruter med jordbruksland i senterpunktet som Landsskogstakseringen har identifisert. Eventuelle varianter vil bestå i å ta et tilleggsutvalg. Alternativ B er å ta et enkelt tilfeldig utvalg av ruter med potensielt jordbruksland og eliminere ruter uten jordbruksland fra utvalget. Varianter vil her foruten ulike mulige definisjoner av ruter med potensielt jordbruksland, bestå i valg av utvalgsstørrelse og eventuell prestratifisering. Fylke og landskapsregion er de aktuelle variablene ved en slik stratifisering.

2.1 Alternativ A: Systematisk utvalg

Utvalgsplanen kan betraktes som en totrinns utvalgsplan. På første trinn trekkes i hver av 8 soner et 3x3-km forband av 1-km²-ruter. Det uttrukne forbandet i hver sone vil vi betrakte som et av 9 mulige forband trukket med sannsynlighet 1/9 og forbandene kan betraktes som at de er trukket uavhengig. Av rutene i disse forbandene er de rutene tatt med i utvalget som har jordbruksland i senterpunktet. Dette betrakter vi som at det er gjort en trekking på annet trinn med trekkesannsynlighet lik antall kvadratkilometer jordbruksland i ruten. Trekkesannsynligheten for par av ruter er ikke like opplagt, men det virker rimelig å bruke en antagelse om uavhengig trekking når varianser skal beregnes. For denne utvalgsplanen foreligger det allerede et utvalg. Landsskogstakseringen har identifisert et utvalg på 1 381 ruter utenom Finnmark. For å fullføre identifiseringen av dette utvalget må senterpunktene i rutene i Finnmark undersøkes. Det ventes å gi omlag 15 ruter til.

Eventuelle variasjoner av utvalgsplanen vil bestå i å supplere dette utvalget. Mest aktuelt er i så fall ekstra ruter i Finnmark, Aust-Agder og fjellområdene, for å få bedre estimater på regionnivå.

2.2 Alternativ B: Enkelt tilfeldig utvalg

Metoden som beskrives her er en videreutvikling av den som ble skissert i det første møtet mellom SSB og NIJOS og omtalt i NIJOS-notatet av 2. februar. På det andre møtet snakket vi om å bruke flere bygningspunkter per bruk og det er dette som legges til grunn her.

Det defineres en populasjon av potensielle jordbruksruter. De potensielle jordbruksrutene bestemmes slik: Landbruksregisteret og GAB kobles for å få en liste over bygninger som tilhører eiendommer med jordbruksareal i drift. GAB har koordinater for disse punktene som brukes til å bestemme hvilken kvadratkilometer rute punktene ligger i. Alle ruter med et slikt bygningspunkt regnes som en potensiell jordbruksrute. Et enkelt tilfeldig utvalg som er noe større enn den ønskede utvalgsstørrelsen trekkes slik at vi får en rangering av rutene i utvalget. (En praktisk måte er å knytte et tilfeldig tall til hver rute med et slikt bygningspunkt og sortere listen etter det tilfeldige tallet.) Det sjekkes om rutene på listen har jordbruksareal, inntil det ønskede antall ruter er kommet med i utvalget. Ruter uten jordbruksareal ekskluderes. Hvis alle ruter med jordbruksareal er kommet med i populasjonen av potensielle jordbruksruter blir resultatet et enkelt tilfeldig utvalg av ruter med jordbruksareal. Et enkelt tilfeldig utvalg innebærer at alle rutene har samme trekkesannsynlighet.

En variant av denne utvalgsplanen er å ta med også naborutene til rutene med bygningspunkter som potensielle jordbruksruter. Vi kan da regne de fire rutene i N, Ø, S og V som har kantlinje mot ruten som naborutene, eller også ta med de fire rutene i NØ, SØ, SV og NV med hjørnepunkt mot ruten.

En annen variasjonsmulighet er å stratifisere rutene etter fylke eller landskapsregion og trekke et enkelt tilfeldig utvalg innenfor hvert stratum på den måten som er beskrevet ovenfor. Trekkesannsynlighetene kan da være ulik i forskjellige strata. Hvis det velges å ta med naboruter, virker det rimelig å ta med naboruter også om de hører til et annet stratum, men å plassere alle ruter i det stratum de selv hører hjemme i.

3. Egenskaper ved utvalgsplanene

I dette avsnittet omtales fordeler og ulemper ved utvalgsplanene og andre kommentarer til disse.

3.1 Egenskaper ved alternativ A

3.1.1 Mulig systematikk i landskapet

I (Schjalm, 1997) ble NIJOS gjort oppmerksom på den generelle risikoen for systematikk i populasjonen som er knyttet til systematisk utvalg. Spesielt vil systematiske underutvalg av 3x3-km forbandene ikke være ønskelig, men det er det heller ikke noe behov for. NIJOS opplyser at det er en avstandsavhengig systematikk i norsk natur med periode omkring 30 km. NIJOS regner det som mindre trolig at skulle finnes en systematikk med periode på 3 km, men vil ikke helt utelukke det. Gårdene i mange norske ligger litt oppe i lia på hver side av dalen, og en kan ikke utelukke at avstanden kan være omtrent 3 km i noen tilfeller.

Vi antar at 3x3-km forband til tross for dette kan være akseptabelt. Selv om det skulle være systematikk i utvalget innenfor hvert dalføre, vil utvalget neppe treffe likt i alle dalfører. Derfor er det grunn til å regne med at de ulike typer områder som kan finnes i et dalføre vil bli dekket av utvalget.

En idé til hvordan effekten av systematisk utvalg kan sjekkes uten noen stor arbeidsinnsats er å lage en mal av papp med vinduer som svarer til et 3x3-km forband. Denne kan legges over et kartblad av det økonomiske kartverket og skyves i forskjellige posisjoner. På et utvalg av kartblader kan det vurderes med øyemål om andelen jordbruksland varierer mye eller lite.

3.1.2 Manglende estimator for førstetrinnsvarians

Et annet problem med det systematiske utvalget er at det ikke tillater forventningsrett estimering av utvalgsvariansen på første trinn. Det foreslås at vi beregner et anslag for denne variansen på grunnlag av en modellbetragtning som uttrykker at det ikke er problemer med periodisitet i landskapet.

3.1.3 Moderat korrelasjon mellom rutene i utvalget

Det systematiske utvalget inneholder ikke naboruter. Da naboruter må antas å være høyere korrelert enn andre ruter, er det sterk grunn til å tro at rutene i et systematisk utvalg er lavere korrelert enn et tilsvarende enkelt tilfeldig utvalg når det gjelder areal av forskjellig type og forekomst av forskjellige fenomener. Med «tilsvarende» tenkes her på kravet om jordbruksland i senterpunktet. Lavere korrelasjon mellom rutene innebærer at utvalget gir mer informasjon.

3.1.4 Dekningen av jordbruksland

Siden sannsynligheten for jordbruksland i senterpunktet øker med mengden jordbruksland i ruten, vil annet trinn i utvalgsplanen resultere i større gjennomsnittlig jordbruksareal per rute i utvalget enn et enkelt tilfeldig utvalg, og derfor gi større datagrunnlag for å estimere fordelingen av jordbruksland enn et enkelt tilfeldig utvalg av samme størrelse. Forskjellen i forventet jordbruksareal per rute er bestemt av fordelingen av jordbruksareal i ruter med jordbruksland.

3.1.5 Fenomener i områder med tynt jordbruksland

På den annen side vil denne utvalgsplanen gi liten dekning av fenomener som spesielt er knyttet til områder med et lavt innslag av jordbruksland. (Om slike fenomener måtte finnes og være av interesse. Kanskje slike ruter vil være av interesse i forbindelse med overvåking av nydyrking?)

En avveining av poengene i 3.1.4 og 3.1.5 vil altså være aktuell ved valg av utvalgsplan.

3.1.6 Ikke behov for å etablere utvalget

Hvis alternativ A velges er det ikke nødvendig å gjøre noe arbeid for å etablere utvalget, idet Landsskogtakseringen har identifisert ruter med jordbruksland i senterpunktet. Det påpekes at dette gjelder ruter som hadde jordbruksland i senterpunktet da Økonomisk Kartverk ble etablert eller ajourført. Tidspunktet for dette varierer, men kan ligge helt tilbake til 1960-tallet. Situasjonen kan derfor være litt annerledes nå, men vi gjør neppe noen vesentlig feil ved å overse dette avviket.

3.1.7 Utvalgsstørrelse

Landsskogtakseringen har identifisert 1 381 ruter utenom Finnmark med jordbruksland i senterpunktet. En tilsvarende gjennomgang for Finnmark ventes å gi omlag 15 ruter til. 2 000 ruter har vært antydnet som NIJOS' kapasitet, men av budsjettmessige grunner er et lavere antall ønskelig. Men det er altså mulig å supplere utvalget med ekstra ruter i regioner hvor det ønskes mer presise estimater enn Landsskogtakseringens utvalg tillater. Finnmark, Aust-Agder og fjellområdene har vært nevnt som aktuelle områder for tilleggsutvalg.

3.2 Egenskaper ved alternativ B

3.2.1 Arbeid med å etablere utvalget

Velges metode B innebærer dette arbeid med å etablere utvalget. Koblingen av GAB og Landbruksregisteret må programmeres, likeså bestemmelsen av ruter fra koordinatene til bygningspunktene. Fra populasjonen av ruter med potensielt jordbruksland må det trekkes et utvalg av kandidatruter, og til slutt må disse undersøkes mot kart for å finne ut om de inneholder jordbruksland. NIJOS anslår arbeidet med koblingen til GAB til et ukeverk.

3.2.2 Valg av strata

Hvis det skal trekkes et stratifisert utvalg må strataene defineres og det må fastsettes en utvalgsstørrelse for hvert stratum. Fylke og landskapsregion er de aktuelle variablene for definisjon av strata.

3.2.3 Variansberegning

For alternativ B kan variansberegninger gjøres med standardmetoder, enten det brukes et enkelt tilfeldig utvalg eller et stratifisert enkelt tilfeldig utvalg.

3.2.4 Dekning av jordbruksland

Alternativ B vil gi et lavere gjennomsnittlig jordbruksareal per rute i utvalget enn alternativ A og følgelig et mindre datagrunnlag for estimering av fordelingen av jordbruksland på forskjellige typer. Forskjellen vil være større hvis naboruter til ruter med bygningspunkter er med i populasjoner enn hvis de ikke er med.

Det virker svært sannsynlig at noe av jordbruksarealet finnes i ruter som ikke har bygningspunkter knyttet til eiendom med jordbruksareal i drift.

3.2.5 Fenomener i områder med tynt jordbruksland

Alt jordbruksland som dekkes av populasjonen av ruter med potensielt jordbruksland vil ha samme trekkesannsynlighet med alternativ B. (Hvis det brukes en stratifisert utvalgsplan gjelder dette for hvert stratum separat, og ikke for hele populasjonen.) Det innebærer at fenomener som spesielt forekommer i områder med et lite innslag av jordbruksland blir bedre dekket enn med alternativ A.

3.3 Sammenlikning av alternativene A og B

3.3.1 Arbeidsmengde

Den samlede arbeidsmengden blir omtrent den samme med de to alternativene. Ved alternativ A må 3x3-rutenettet for Finnmark gjennomgås, noe NIJOS anslår til et månedsverk, men dette arbeidet trenger ikke gjøres før i år 2000. For resten av landet kan et ferdig utvalg hentes frem. Alternativ B krever en kobling mellom Landbruksregisteret og GAB, noe NIJOS anslår til et ukeverk. Dette arbeidet kan innebære en viss risiko for forsinkelser i en tidskritisk fase av prosjektet. Det betyr at utvalget er raskere klart og at faren for forsinkelser i denne fasen elimineres eller reduseres med alternativ A.

3.3.2 Dekning

Alternativ A gir bedre dekning av jordbruksland enn alternativ B (med samme utvalgsstørrelse), men dårligere dekning av tynt jordbruksland.

3.3.3 Variansberegning

Alternativ B tillater standardmetoder for variansberegninger, mens alternativ A gir behov for å utvikle en variansberegningmetode for første trinn. Denne metoden kan ikke bli designforventningsrett.

4. Estimering

Ved alternativ A vil kilene (ruter langs sonegrensene som er mindre enn full størrelse) i UTM-33 representere et kompliserende element ved estimeringen fordi disse rutene er ekskludert fra trekkingen. I den grad kilene ikke kan komme med i utvalget er det naturlig å kompensere dette ved å gi naborutene til kilene høyere vekt. Dette vil gjøre estimerings- og variansberegningformlene noe mer kompliserte.

Alternativ B tillater derimot alle områder å komme med i utvalget.

En enkel idé er å bruke Horvitz-Thompson estimatoren for totaltall, og plugge inn disse estimatene for parametere som kan uttrykkes som funksjoner av totaltall. En annen mulighet er å kombinere informasjon fra utvalget med registerdata for alle ruter. Slik informasjon kan utnyttes ved å bruke en etterstratifiseringsestimator eller regresjonsestimator. Det må da velges ut registervariabler som antas

å være forholdsvis sterkt korrelerte med interessevariablene. En slik estimeringsteknikk innebærer at det gjøres modellantakelser, men det finnes variansformler også for disse situasjonene.

5. Utvalgsstørrelse

Utvalgsstørrelse kan ta utgangspunkt i budsjettet eller i en ønsket presisjon. Beregning av utvalgsstørrelse ut fra ønsket presisjon krever informasjon om interessevariablenes fordeling i populasjonen, noe vi bare kan gjøre gjetninger om før dataene foreligger. Vi kan eventuelt bruke data fra tidligere liknende undersøkelser. For andeler av ja/nei-variable kan det beregnes utvalgsstørrelser som er tilstrekkelige til å sikre en gitt presisjon. (Her er sett bort fra problemet med førstetrinnsvariansen i alternativ A.).

Ved bruk av Horvitz-Thompson estimatoren gir et utvalg på 1 500 ruter en presisjon for andelen av ruter i landet som har et gitt kjennemerke på linje med de fleste meningsmålinger, siden disse har omtrent denne utvalgsstørrelsen. For regiontall blir presisjonen lavere. Vi antar at presisjonen både for landstall og regiontall kan bedres ved å utnytte registeropplysninger. Det må da gjøres modellantakelser.

For beregning av utvalgsstørrelse fra budsjettet kan settes opp følgende betingelser:

La n_F være antall ruter som fotograferes, n_T antall ruter som tolkes, α variable fotograferingskostnader per rute (film, flytid etc.), β tolkningskostnader per rute, c_0 faste kostnader, c totale kostnader og B budsjettet.

Samlede kostnader blir $c = \alpha n_F + \beta n_T + c_0$.

Budsjettbetingelsen er $c \leq B$.

Dessuten må alle ruter som tolkes også fotograferes: $n_T \leq n_F$.

Fra disse betingelsene følger at $n_T \leq \frac{B - c_0}{\alpha + \beta}$, dvs. at øvre grense blir forholdet mellom tilgjengelige midler til variable kostnader og samlede variable kostnader per rute som fotograferes og tolkes.

Hvis vi velger $n_F > \frac{B - c_0}{\alpha + \beta}$, blir det ikke råd til å tolke alle rutene som fotograferes, og vi får en lavere grense for n_T , nemlig $n_T \leq \frac{B - c_0 - \alpha n_F}{\beta}$. Et slikt valg av n_F gir mulighet for å estimere varianser fra deler av utvalget og øke utvalgsstørrelsen litt i strata med høy varians. Det innebærer imidlertid at den totale utvalgsstørrelsen blir mindre enn hvis vi setter $n_T = n_F = \frac{B - c_0}{\alpha + \beta}$ og at det fotograferes ruter som ikke tolkes. Antagelig er det fornuftig å la $n_T = n_F$. Å la $n_T < n_F$ er mest aktuelt hvis variansene er svært ulike i forskjellige strata, men at vi har lite forhåndskunnskap om disse variansene.

Herfra forutsettes at $n_T = n_F = n$.

Ved alternativ B må det fastsettes en samlet utvalgsstørrelse og en fordeling på strataene. Ved denne fordelingen kan det bli en avveining mellom et ønske om best mulige estimater for Norge og gode

estimerer for alle strata/regioner. Hensynet til det siste kan tilsi et høyere antall ruter i områder med få ruter på bekostning av de andre.

I basisversjonen av alternativ A er utvalgsstørrelsen (og også utvalget) fiksert. Utvalgsstørrelsen blir svært nær 1 400. For et eventuelt tilleggsutvalg må de aktuelle regionene det skal være tilleggsutvalg i bestemmes. Det må videre bestemmes om tilleggsutvalget skal trekkes fra alle ruter i disse regionene som ikke er med i det fikserte utvalget, fra ruter med jordbruksland, eller fra ruter med jordbruksland i senterpunktet. Ruter med jordbruksland i senterpunktet synes mest nærliggende; Det gir en økt utvalgsstørrelse, men ellers et utvalg etter samme prinsipper som det fikserte utvalget. Hvis hensikten med tilleggsutvalget er å bedre deknningen av tynt jordbruksland kan imidlertid et utvalg av ruter med jordbruksland være aktuelt.

Vårt forslag til utvalgsplan for eventuelle tilleggsruter er å bruke et enkelt tilfeldig utvalg i hvert stratum/region av aktuelle ruter, dvs. av ruter i stratumet som ikke er med i det fikserte utvalget og som tilfredsstillende kriteriet for tilleggsrutene.

Til slutt må det fastsettes et antall tilleggsruter for hvert stratum.

6. Anbefalinger

Vi anbefaler at det systematiske utvalget benyttes, altså alternativ A for utvalgsplan. Ved estimeringen bør det gjøres modellantakelser, slik at registeropplysninger om hele landbruksarealet kan utnyttes. Disse opplysningene utnyttes ved å bruke en etterstratifiseringsestimator eller eventuelt en regresjonsestimator.

Litteratur:

Geir-Harald Strand: *Utvalgs- og Beregningsmetoder for Kulturlandskapsobservasjon*. NIJOS, 21. april 1997.

Arnfinn Schjalm: *Notat om Kulturlandskapsobservasjon*. SSB, 16. mai 1997.

De sist utgitte publikasjonene i serien Notater

- 98/77 M.H. Erichsen og T. Halvorsen: Marshall-planen og norsk offisiell statistikk. 20s.
- 98/78 K.A. Brekke og R. Aaberge: Ekvivalensskala og velferd. 18s.
- 98/79 E. Gulløy, S. Opdahl og I. Øyangen: Levekår og forbruk blant studenter 1998: Hovedresultater og dokumentasjon. 174s.
- 98/80 D. Roll-Hansen: Forbruksundersøkelsen 1997: Dokumentasjonsrapport. 88s.
- 98/81 Ø. Døhl: Temperaturkorrigering av energiforbruket: En empirisk analyse. 109s.
- 98/82 T. Vogt: Dokumentasjonsrapport AKU - 1997. 44s.
- 98/83 A.B. Svinnet: Plan for SSBs arbeid med KOSTRA fram til fullskala drift. 25s.
- 98/84 D. Roll-Hansen, L. Solheim og L.C. Zhang: Kopiering ved universiteter og høyskoler. Korrigert utgave. 88s.
- 98/85 T. Vogt: Dokumentasjonsrapport - Arbeidsmiljø og omstilling. 18s.
- 98/86 A. Hallenstvedt og E. Sørensen: Omsetningsstatistikk for industrien. 29s.
- 98/87 B. Mathisen: Flyktninger og arbeidsmarkedet 4. kvartal 1997. 40s.
- 98/88 J. Sexton: Fremskrivning av tidsserier i KNR. 20s.
- 98/89 A.H. Foss: Definisjoner og beregningsmetoder for dødelighetstabell. 16s.
- 98/90 T. Dale: Samordnet levekårsundersøkelse 1998 - panelundersøkelsen: Dokumentasjonsrapport. 95s.
- 98/91 L. Lindholt: Rammevilkår for energigjenvinning av plast. 14s.
- 98/92 T. Vogt: Folatkunnskap blant kvinner i fertil alder: Dokumentasjonsrapport. 17s.
- 98/93 B.L. Western: Beregning av vekter til inntekts- og formuesundersøkelsene 1995. 14s.
- 98/95 R. Johannessen: Prisindeks for hotellovernatting - delundersøkelse i konsumprisindeksen. 18s.
- 98/96 K.J. Einarsen: Definisjonskatalog for videregående opplæring: Utarbeidet av arbeidsutvalget i FylkesKOSTRA-utdanning og statistikkgruppen i Prosjekt LINDA-opplæring. 1. Utgave. 27s.
- 98/97 K.A. Brekke: Om metoder for beregning av miljøprofil for ulike varer, og hva vi trenger det til. 20s.
- 98/98 I.S. Wold: Modellering av husholdningenes transportkonsum for en analyse av grønne skatter: Muligheter og problemer innenfor rammen av en nyttetremodell. 81s.
- 98/99 R. Gudem: Utvikling av statistikk over bygg- og anleggsavfall. 56s.
- 98/100 T. Skjerpen: Konsumfordelingssystemet i KVARTS: Teknisk dokumentasjon. 42s.
- 99/1 F. Thorkildsen: FoB2000. Kobling av adresser fra GAB og DFS - status og utvikling. 22s.
- 99/2 K. Ibenholt og K.A. Brekke: Rammevilkår for produksjon av brunt papir. 12s.
- 99/3 I. Øyangen: Liv og helse i Akershus: Dokumentasjonsrapport. 22s.
- 99/4 M.V. Dysterud og E. Engelian: Tettstedsavgrensning 1998: Teknisk dokumentasjon av nye rutiner. 53s.
- 99/5 Samordnede levekårsundersøkelser: Rapport fra brukerseminar 22. april 1996 og høringsuttalelser. 46s.
- 99/6 H. Utne: FoB2000. Sektordata: Statusrapport og tiltaksplan. 86s.
- 99/7 M. Sjøberg: Instruksjonar til og data frå eksperiment om internasjonal kvotehandel. 27s.

Notater



Tillatelse nr.
159 000/502

B *Returadresse:*
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Statistisk sentralbyrå

Oslo:
Postboks 8131 Dep.
0033 Oslo

Telefon: 22 86 45 00
Telefaks: 22 86 49 73

Kongsvinger:
Postboks 1260
2201 Kongsvinger

Telefon: 62 88 50 00
Telefaks: 62 88 50 30

ISSN 0806-3745



Statistisk sentralbyrå
Statistics Norway