

# Interne notat

STATISTISK SENTRALBYRÅ

86/14

13. februar 1986

## HELSE OG FRITIDSAKTIVITET, YRKE OG INNTEKT \*)

Ein metodedokumentasjon

av

Asbjørn Torvanger

### INNHALD

	Side	
I	INNLEIING	2
II	VAL AV METODE	3
III	SAMANHENGEN MELLOM HELSE OG FRITIDSAKTIVITET, YRKE OG INNTEKT	
	1. Data	4
	2. Variablane for helse som inngår	4
	3. Den Samansette Helseindikatoren (SHI)	5
	4. Dei uavhengige variablane	9
	5. Modellane	10
	6. Regresjonane; problem med metoden	12
	7. Kanonisk korrelasjonsanalyse	14
IV	RESULTAT	
	1. Forklaringskrafta til modellane	15
	2. Korrelasjon mellom dei uavhengige variablane	15
	3. Hovudresultat frå analysen	16
	4. Resultat frå den kanoniske korrelasjons- analysen	18
	5. Oppsummering	19
V	ERFARINGAR MED METODEN	19
	APPENDIKS Variablane som inngår i analysen	22
	REFERANSAR	26

\*) Inngår i prosjektet 'Miljø og Levekår'

## I INNLEIING

Seksjon for ressurs- og miljøanalyser ved Statistisk Sentralbyrå har i eit par år arbeidd med prosjektet 'Miljø og Levekår'. I eit av delprosjekta, 'Friluftaktivitet og Helse', blir samanhengen mellom friluftaktivitet og helse studert.

Analysen som blir dokumentert i dette notatet inngår som ein del av 'Friluftaktivitet og Helse'. Utgangspunktet for analysen er følgjande hypotese:

Aukande fritidsaktivitet og aukande inntekt har ein positiv samanheng med helsetilstanden til folk. Også yrke og fysisk aktivitet i arbeidet er viktig for helsetilstanden.

Analysen har brukt data som alt er samla inn i Statistisk Sentralbyrå. Standard minste kvadrat regresjon vart valt som analysemetode, sjølv om dei fleste tilgjengelege data er kategoriske. I eit anna delprosjekt som studerer samanhengen mellom støy i bustadmiljø og psykisk helse, vart log-lineær analyse nytta. Ved log-lineær analyse kan ein studere i kva grad variablar i fleir-dimensjonale tabellar avheng av kvarandre, og metoden er såleis betre tilpassa kategoriske variablar.

Ei oppsummering av resultatata frå delprosjektet 'Friluftaktivitet og Helse' kan finnast i sluttrapporten frå hovudprosjektet 'Miljø og Levekår' (SSB: Miljø og Levekår, Rapport 1986/-).

For å studere samanhengen mellom helse og friluftaktivitet, yrke og inntekt bør ein livsløps-analyse gjennomførast, og til dette trengst det tidsseriar av individ-data. Desverre finst ikkje slike dataseriar, så denne analysen måtte ta utgangspunkt i observasjonar frå berre eit tidspunkt. Dersom verdiane på variablane er relativt stabile over tid for kvar person, vil ein observasjon i brukbar grad representere ein tidsserie. Ein slik føresetnad ligg bak denne analysen. Det er også sett bort frå at det kan ta tid før ei endring i dei uavhengige variablane slår ut i endra helsetilstand. Desse føresetnadane er relativt sterke. For fleire av variablane som inngår i analysen vil verdien vanlegvis avhenge av alderen til ein person.

Det viste seg at modellane i liten grad var i stand til å forklare

variansen i datamaterialet. Den viktigaste grunnen til dette er sannsynlegvis at metoden som er nytta ikkje passer så godt til analyse med kategoriske variablar. Den samansette helseindikatoren SHI passer heller ikkje så godt til metoden når variablane som inngår er kategoriske. Eit anna viktig problem er at testing av hypotesen krev ein livsløps-analyse, medan det berre finst observasjonar frå eit tidspunkt i data-materialet.

Resultata gjev ein indikasjon på støtte til hypotesen om at helsetilstanden til folk har ein positiv samanheng med aukande fritidsaktivitet (mosjon), men ingen indikasjon på støtte til hypotesen om ein positiv samanheng mellom helse og inntekt.

Framover kan ein enten satse på å bruke dei (kategoriske) data som finst og drive metodeutvikling, eller planlegge integrerte prosjekt der hypotese, data-innsamling (datakvalitet) og metode blir tilpassa kvarandre.

## II VAL AV METODE

I dette delprosjektet er det brukt regresjonsanalyse på data frå Helseundersøkelsen 1975, og desse er stort sett kategoriske (dvs. at variablane blir tilordna verdiar som plasserer teljeeiningane i ulike kategoriar; variablane er diskrete). Regresjonsanalyse vart valt fordi metoden er enkel å bruke og gjev resultat som er relativt lette å tolke, sjølv om metoden er mindre eigna for kategoriske variablar. Også til dømes log-lineær analyse har sine svake sider, blant anna er resultata vanskelegare å tolke enn resultata frå ein regresjonsanalyse.

Ved å gjennomføre regresjonsanalyser kunne ein samansett helseindikator prøvast (sjå III.3). Denne vart konstruert ved å kombinere åtte helse-variablar. Dei variablane som inngår kunne dermed samla overførast til intervall-nivå. Utgangspunktet for ein regresjonsanalyse ville med dette bli betre. Regresjonsanalysane er supplert med ein kanonisk korrelasjonsanalyse (sjå III.7).

### III SAMANHENGEN MELLOM HELSE OG FRITIDSAKTIVITET, YRKE OG INNTEKT

#### 1. Data

Analysen byggjer på Helseundersøkelsen 1975. Også Friluftundersøkelsen 1974, Friluftundersøkelsen 1970, Levekårsundersøkelsen 1983 og Levekårsundersøkelsen 1980 vart vurderte. Helseundersøkelsen 1975 vart valt fordi dette er den einaste datakjelde med både helse-data og fritidsaktivitet-data.

Helseundersøkelsen 1975 er basert på intervju av eit netto-utval på 11.014 personar og vart gjennomført i oktober 1975. Utvalet er trekt frå alle personar busette i landet utanom dei som oppheldt seg på helseinstitusjon, aldersheim eller barneheim. I dei uttrekte hushalda vart alle personar intervjuet. Utvalet skal gje eit representativt tverrsnitt av folkesetnaden i Noreg. Formålet var å gje eit betre oversyn over sjukdomar og skader blant folk utanfor helseinstitusjonar, og å undersøke i kva grad sjukdom eller skader fører til nedsett aktivitet. Ein ville også kartlegge kontakt med helsevesenet, bruken av helsetenester og kva konsekvensar sjukdom og skader har for den sjuke sjølv og familien.

Analysen bruker altså berre observasjonar frå eit tidspunkt. Dessutan gjeld mange av spørsmåla helse-situasjonen dei siste 14 dagar. Hadde det same utvalet blitt intervjuet fleire gonger med eit visst mellomrom, dvs. om panel-data var tilgjengeleg, kunne analysen ha gjeve sikrere resultat.

#### 2. Variablane for helse som inngår

I analysen er 'Helse' avhengig variabel. I Helseundersøkelsen 1975 inngår det mange variablar for helse. Åtte variablar vart plukka ut etter ei vurdering av kva variablar som best reflekterte folk sin helsetilstand. Denne vurderinga vart gjort i samråd med ein medisinar. Variablane er:

- \* Nedsett yrkesaktivitet pga sjukdom eller skade; ja/nei
- \* Fråver frå arbeidet siste 14 dagar pga sjukdom eller skader; målt i talet på dagar

- \* Kor lenge sidan siste kontakt med lækjar i eller utanfor institusjon; 7 kategoriar frå under 2 veker til aldri
- \* Fysisk evne ('førlighet'); 5 kategoriar frå ingen fysisk evne til gå i trapp/gå 15 min/springe 50 m
- \* Kontakt med helsetenesten; summen av 11 variablar som kvar gjev gonger kontakt med helsetenesten siste 14 dagar
- \* Kontakt med helsetenesten av meir alvorleg karakter; summen av 3 variablar som kvar viser gonger kontakt med helsetenesten av meir alvorleg karakter siste 14 dagar
- \* Har lækjar sagt at personen har hatt nærvøse/psykiske plager?; ja/nei
- \* Har personen brukt roande middel eller nervemedisin?; ja/nei

Appendikset gjev fleire detaljar om desse variablane. Variablane gjev ei nokså grov klassifisering av helsetilstanden til folk.

### 3. Den Samansette Helseindikatoren (SHI)

Dei åtte variablane i forrige avsnitt dekker berre nokre sider av helsetilstanden til ein person. Difor vart det prøvd å lage eit samla mål på helsetilstand, dvs. ein samansett helseindikator. Denne er samansett av dei åtte helse-variablane i forrige avsnitt. Kvar klasse av dei kategoriske variablane blei tilordna ei vekt ut frå ei vurdering av 'avstanden' i helsetilstand mellom kvar kategori. Vektene er spesifiserte i Appendikset.

Variablane vart standardiserte (ved å trekke frå det aritmetiske gjennomsnittet og dele på standardavviket) for å sikre at kvar av variablane skulle ha like stor vekt i SHI. Dei variablane som berre har verdien 0 eller 1 (ja eller nei) vart standardiserte ved:

$$(1) \quad ZX_i = \frac{X_i - \bar{X}}{\sqrt{\bar{X}(1-\bar{X})}}$$

der  $X_i$  er den opprinnelege variabelen og  $\bar{X}$  det aritmetiske gjennomsnittet.

Utan standardisering ville skaleringa av kvar variabel kunne slå tilfeldig ut i SHI. Ein variabel med stor avstand mellom øvre og nedre klasse ville telje meir for SHI enn ein annan variabel med liten avstand. Dei nye variablane blir lik avviket frå gjennomsnittet målt i

standardavvik, og vil ha forventning 0 og standardavvik lik 1. Kvar av variablane i SHI vart splitta opp etter alder/kjønns-gruppe og standardiserte innan gruppa, dvs. ved å bruke gjennomsnittet og standardavviket innan gruppa. Deretter kunne SHI innan kvar gruppe reknast ut ved å summere/subtrahere dei standardiserte variablane.

SHI vart rekna ut ved å addere/subtrahere dei åtte standardiserte variablane. Ein person med betre helsetilstand enn ein annan vil få høgare verdi på SHI. For aldersgruppa 6-14 år vart 'Fråver frå arbeidet' og 'Fysisk evne' teke ut av SHI då dei ikkje er registrerte. Formelen til SHI er vist i appendikset.

Diagram 1 viser fordelinga til SHI i desilar for utvalet samla, Diagram 2 for menn og kvinner 6-14 år, Diagram 3 for menn og kvinner 15-54 år og Diagram 4 for menn og kvinner som er 55 år og eldre. Dette er ei kumulativ fordeling der personane er ordna etter stigande verdi på SHI. Desil 1 viser SHI-verdien til ein person som er plassert slik at 10% av utvalet har lavare SHI-verdi enn han/ho. Tilsvarande viser desil 2 SHI-verdien til ein person som er plassert slik at 20% av utvalet har lavare SHI-verdi enn han/ho osv.

Diagram 1 Fordelinga til SHI for utvalet samla (i desilar)

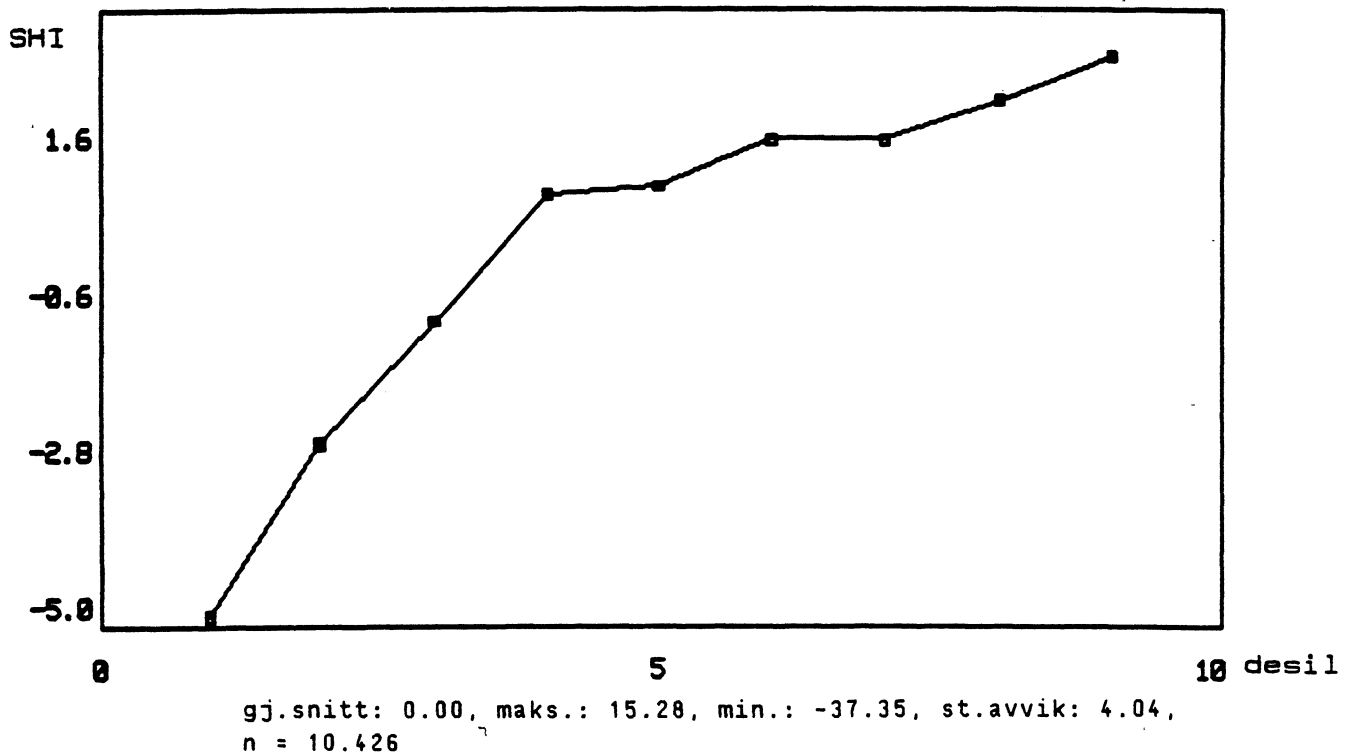
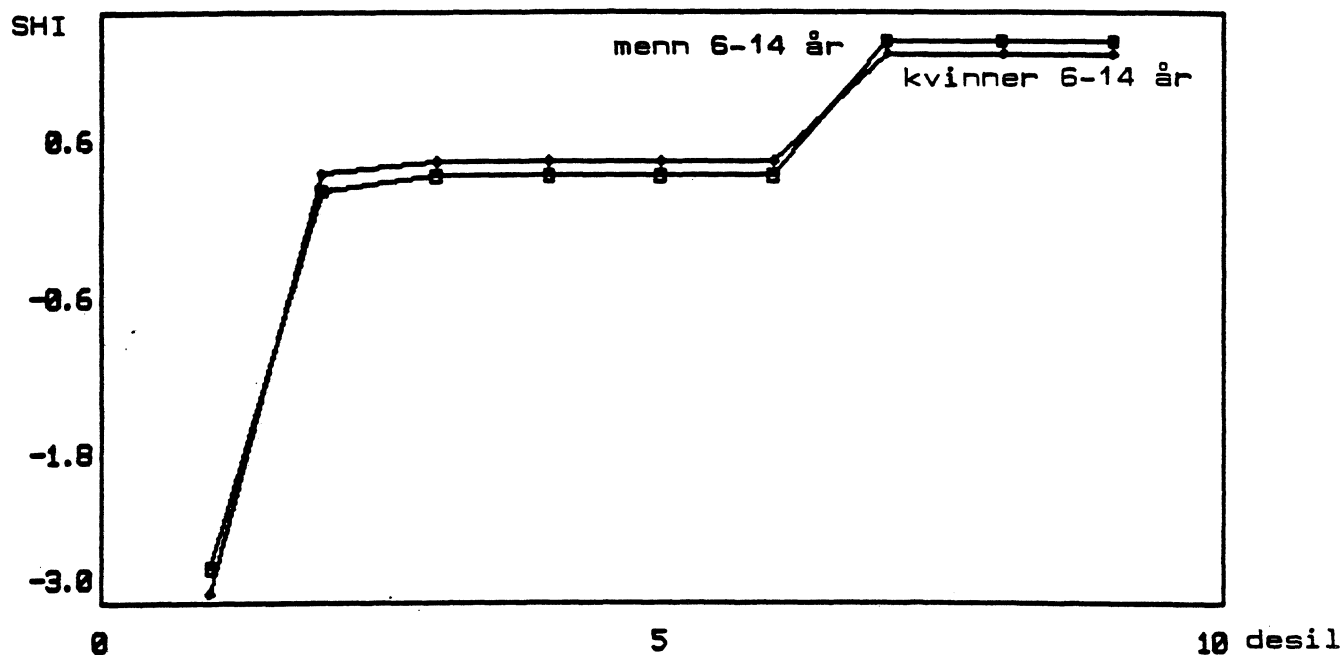
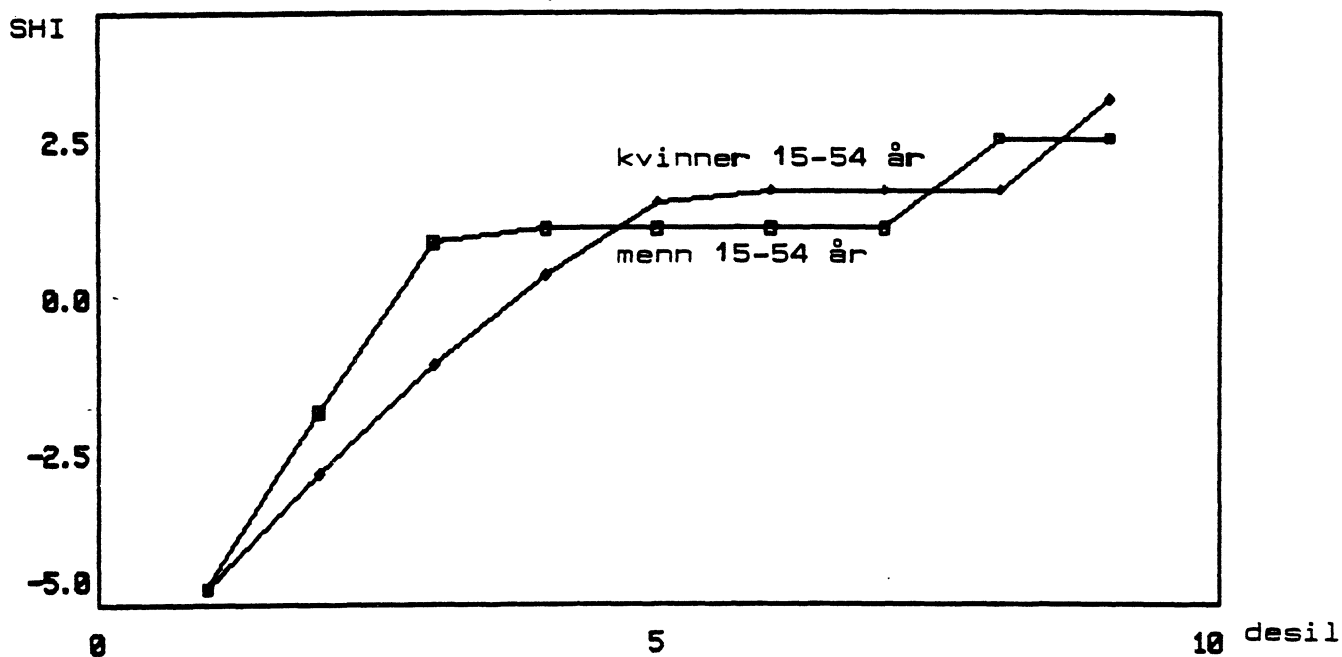


Diagram 2 Fordelinga til SHI for menn og kvinner 6-14 år (i desilar)



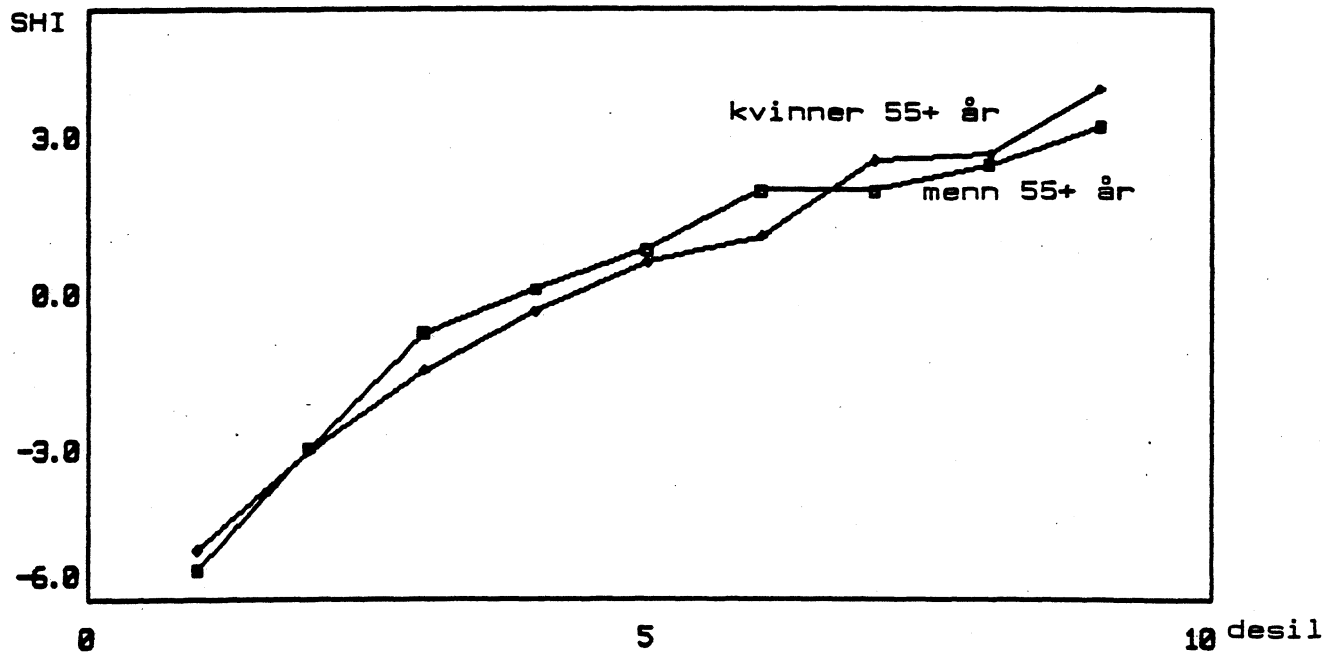
Menn 6-14 år: gj.snitt: -0.00, maks.: 12.52, min.: -22.19,  
 st.avvik: 2.93, n = 884  
 Kvinner 6-14 år: gj.snitt: 0.01, maks.: 10.28, min.: -26.70,  
 st.avvik: 2.95, n = 799

Diagram 3 Fordelinga til SHI for menn og kvinner 15-54 år (i desilar)



Menn 15-54 år: gj.snitt: -0.00, maks.: 17.81, min.: -34.36,  
 st.avvik: 4.16, n = 2.639  
 Kvinner 15-54 år: gj.snitt: 0.00, maks.: 19.01, min.: -27.74,  
 st.avvik: 4.08, n = 2.722

Diagram 4 Fordelinga til SHI for menn og kvinner 55 år og eldre (i desilar)



Menn 55+ år: gj.snitt: -0.00, maks.: 13.84, min.: -26.90,  
st.avvik: 4.08, n = 1.211  
Kvinner 55+ år: gj.snitt: 0.00, maks.: 15.92, min.: -21.60,  
st.avvik: 3.99, n = 1.273

Dersom ei kurve har jamn stigning betyr det at SHI skiller godt mellom personane i gruppa. Det går fram at SHI i liten grad skiller mellom personar i den yngste aldersgruppa, både for jenter og gutar. SHI skiller best mellom personar i den eldste aldersgruppa. Det er liten skilnad på kvinner og menn for denne gruppa. I aldersgruppa 15-54 år er SHI best for kvinner. For utvalet samla viser Diagram 1 at SHI skiller bra.

Kurvane tyder ikkje på at SHI er normalfordelt. For mange personar ligg rundt gjennomsnittet for SHI. Dessutan er venstre hale av fordelinga (dvs dei som har dårlegare helsetilstand enn gjennomsnittet - har negativ verdi på SHI), lenger enn høgre hale (-37.4 mot 15.3 for utvalet samla). Dette tyder på at SHI stort sett måler negative avvik frå gjennomsnittleg helsetilstand.

Som ein kunne vente på grunn av standardiseringa ligg det aritmetiske gjennomsnittet til SHI rundt null. Standardavviket til SHI ligg i underkanten av 3 for den yngste aldersgruppa medan det i utvalet samla og for dei andre aldersgruppene ligg rundt 4. Dette viser igjen at SHI skiller dårlegast mellom individ i den yngste aldersgruppa.



#### 4. Dei uavhengige variablane

Dei viktigaste uavhengige variablane i analysen er 'Fritidsaktivitet', 'Inntekt' og 'Yrke'. 'Yrke' er inndelt i to variablar: 'Yrkesfelt' og 'Fysisk aktivitet i arbeidet'. I tillegg er nokre andre variablar som har mykje å seie for folk si helse teke med: 'Kjønn', 'Alder', 'Røykevanar', 'Alkoholbruk', 'Koronar hjarteskjukdom', og 'Region'. Dermed er desse variablane med:

- \* Fritidsaktivitet siste 12 månader; inndelt i klassane: set i ro, lett mosjon meir enn 4 timar i veka, tung mosjon meir enn 4 timar i veka, og trener hardt
- \* Brutto hushaldsinntekt pr forbrukseining ifølgje skattelikninga; rekna i 100 1975 kroner
- \* Yrkesfelt og type aktivitet; 10 klassar
- \* Fysisk aktivitet i arbeidet; inndelt i klassane: set i ro i arbeidet, står og går mykje i arbeidet, går og løfter mykje i arbeidet, og tungt kroppsarbeid
- \* Kjønn
- \* Alder; inndelt i aldersgruppene: 6-14 år, 15-54 år, og 55 år og eldre
- \* Røykevanar; inndelt i klassane: dagleg røykar med 15 eller fleire sigarettar, dagleg røykar med 10-14 sigarettar, dagleg røykar med 1-9 sigarettar, før dagleg røykar, og aldri røykt
- \* Alkoholbruk; inndelt i klassane: bruker ikkje alkohol, bruker alkohol svært sjeldan, bruker alkohol av og til, og bruker alkohol nesten dagleg
- \* Koronar hjarteskjukdom; har eller har hatt ein av hjarteskjukdomane i denne gruppa
- \* Region; landet er inndelt i 11 regionar

Fleire detaljar kan finnast i appendikset. Rørslehemma vart skilt ut frå analysen då dei skiller seg vesentleg ut frå andre grupper når det gjeld helsetilstand. Definisjonen på rørslehemma er dei som fell i kategorien 'ingen funksjonsevne ('ferdighet')' til variabelen 'Behov for dagleg hjelp'. Personar med inntekt lik null er ikkje teke med i analysen, då inntekt lik null skaper problem i ein av modellane. Gruppa med inntekt lik null er relativ liten. Aldersinndelinga er grov, men det var nødvendig å redusere talet på separate regresjonar. Venteleg er skillet over eller under 55 år viktigast når det gjeld den almenne

helsetilstanden til folk. Ei meir utfyllande aldersinndeling kunne vere: 6-14 år, 15-34 år, 35-54 år, 55-74 år, og 75 år og eldre.

### 5. Modellane

To ulike modellar blir prøvd; ein modell der variablane inngår multiplikativt (dei uavhengige variablane er multiplisert med kvarandre) og ein modell der variablane inngår additivt (dei uavhengige variablane er adderte). Med utgangspunkt i datamaterialet som skal brukast, er modellane gjort så enkle som mogeleg.

I den multiplikative modellen MODMULT er den marginale effekten av ei endring i ein av dei uavhengige variablane, dvs. effekten på SHI av ei lita endring i variabelen, avhengig av nivået på dei andre uavhengige variablane.

I den additive modellen MODADD er den marginale effekten av ei endring i ein av dei uavhengige variablane uavhengig av nivået på dei andre uavhengige variablane.

I begge modellane inngår inntekta som kontinuerleg variabel, medan alle andre variablar er kategoriske og inngår som binære variablar (dummy-variablar). Den additive modellen er enklare å handtere matematisk. Den er også enklare å tolke. Den multiplikative modellen må omformast (ved å ta logaritmen av den) for å kunne estimerast med enkle metodar, og dette fører til problem dersom ein variabel kan vere negativ eller null.

#### MODMULT

$$(2) \text{SHI}_i = c I_i^b e^{\sum_{j=1,3,4} f_j \text{DF}_{ji}} e^{\sum_{j=2}^4 c_j \text{DB}_{ji}} e^{\sum_{j=1}^4 d_j \text{DS}_{ji}} e^{\sum_{j=2}^4 a_j \text{DA}_{ji}} e^{\sum_{j=2}^{10} y_j \text{DY}_{ji}} e^{kDK_i} e^{\sum_{j=2}^{11} r_j \text{DR}_{ji}} e^{u_i}$$

Ved å ta logaritmen til (2) får vi:

$$(3) \ln \text{SHI}_i = \ln c + b \ln I_i + \sum_{j=1,3,4} f_j \text{DF}_{ji} + \sum_{j=2}^4 c_j \text{DB}_{ji} + \sum_{j=1}^4 d_j \text{DS}_{ji} + \sum_{j=2}^4 a_j \text{DA}_{ji} + \sum_{j=2}^{10} y_j \text{DY}_{ji} + kDK_i + \sum_{j=2}^{11} r_j \text{DR}_{ji} + u_i$$

MODADD

$$(4) \text{SHI}_i = A + vI_i + \sum_{j=1,3,4} f_j \text{DF}_{ji} + \sum_{j=2}^4 c_j \text{DB}_{ji} + \sum_{j=1}^4 d_j \text{DS}_{ji} \\ + \sum_{j=2}^4 a_j \text{DA}_{ji} + \sum_{j=2}^{10} y_j \text{DY}_{ji} + k\text{DK}_i + \sum_{j=2}^{11} r_j \text{DR}_{ji} + w_i$$

der:

- i : indeks for person/observasjon
- j : dummy-variabel indeks; sjå Appendiks
- SHI : samansett helseindikator
- C, A : konstantar
- $u_i, w_i$  : stokastiske restledd
- e : konstanten e, basis for naturlege logaritmer
- b : koeffisient som gjev prosentvis endring i SHI ved 1% endring i Inntekt
- v : koeffisient som viser marginal effekt på SHI ved endring i Inntekt
- $f_j$  : dummy-variabel koeffisient for fritidsaktivitet
- $c_j$  : dummy-variabel koeffisient for fysisk aktivitet i arbeidet
- $d_j$  : dummy-variabel koeffisient for røykevaner/sigarettforbruk
- $a_j$  : dummy-variabel koeffisient for alkoholforbruk
- $y_j$  : dummy-variabel koeffisient for yrke/type aktivitet
- k : dummy-variabel koeffisient for koronar hjertesjukdom
- $r_j$  : dummy-variabel koeffisient for region

Den stokastiske strukturen er representert med restledda  $u_i$  og  $w_i$ . Dei måler uspesifiserte faktorar sin effekt på SHI. Det blir rekna at restledda er normalfordelte med forventning lik null og konstant varians, og at dei er ukorrelerte. I avsnitt III.3 går det fram at SHI ikkje er normalfordelt og dermed er heller ikkje restledda normalfordelte. Sidan utvalet er stort kan ein likevel som ei tilnærming rekne at restledda er normalfordelte. Det kan vere ein tendens til at personar med lav verdi på SHI har relativt større utslag på dei uavhengige variablane. I så fall vil variansen stige når SHI minkar. Til dømes kan det tenkast at personar med dårleg helse (dvs. lav verdi på SHI) har ei spesielt lav inntekt. Det er mindre fare for at restledda er korrelerte fordi analysen ikkje bruker tidsseriar.

Dummy-variabel koeffisientane har same symbol i begge modellane for å forenkle framstillinga. I realiteten blir ulike koeffisientar estimert ettersom modellane er ulike.

## 6. Regresjonane: problem med metoden

Det vart utført standard minste kvadrat regresjon på MODMULT ved likning (3) og på MODADD ved likning (4). Regresjonane vart gjennomført for utvalet samla og for kvar dei seks alder/kjønns-gruppene. Statistikk-pakken SPSS-X blei brukt. For MODMULT blir  $\ln C$  estimert som konstant og  $b$  som koeffisienten til  $\ln I$ . Dummy-koeffisientane  $f_j, c_j, d_j, a_j, y_j, k$  og  $r_j$  blir som følgje av metoden estimerte som skift i konstanten  $\ln C$ . For MODADD blir  $A$  estimert som konstant og  $v$  som koeffisienten til  $I$ . Dummy-koeffisientane blir estimerte på same måte som for MODMULT.

Fordi logaritmen til eit negativt tal ikkje er definert, blir det problem med MODMULT og likning (3) for negative verdiar av SHI. Problemet vart løyst ved å legge 100 til verdien av SHI før logaritmen vart rekna ut. Denne transformasjonen skulle berre ha noko å seie for tolkninga av dei estimerte koeffisientane med omsyn på skala.

For ein del personar er spørsmål ikkje stilt eller svar manglar. Desse svar-klassane er handsama særskilt i analysen. Fordi rørslehemma skiller seg sterkt ut frå andre når det gjeld helsetilstand, prøver ein å unngå å ta denne gruppa med i analysen. Dette vart gjort ved ikkje å ta med personar som fell i kategorien 'Ingen fysisk evne ('førlighet')' til helse-variabelen 'Behov for dagleg hjelp'. Dermed blir rørslehemma definert som personar som ikkje er i stand til å utføre personleg hygiene, av- og påkledning og daglege innkjøp. Dei som ikkje har svart på spørsmålet eller spørsmålet ikkje er stilt til, blir ikkje rekna som rørslehemma, og er difor teke med. Det er 558 personar som enten er rørslehemma eller har null i inntekt (sjå III.4). Når det gjeld helse-variabelen 'Fysisk evne ('førlighet')' er gjennomsnittet for kvar alder/kjønns-gruppe sett inn dersom svar manglar. I analysen av heile utvalet er gjennomsnittet for dette sett inn når svar manglar. For dei andre helse-variablane manglar så få observasjonar at det ikkje vart gjort noko spesielt med desse.

I kvar regresjonsanalyse blir korrelasjonskoeffisienten mellom to variablar rekna ut ved å ta med alle personar som har registrert ein verdi på desse to variablane. Om ein person manglar verdi på andre variablar blir han/ho likevel teke med.

Det er sannsynleg at nokre av dei uavhengige variablane er korrelerte, til dømes yrke og inntekt. Dette kan gje skeive estimat i regresjonane fordi det blir vanskeleg å identifisere effekten av den einaskilde variabel. Problemet kan reduserast ved å ta ut av regresjonen alle variablar som er sterkt korrelerte utanom ein. Fordi vi mangla tilstrekkeleg informasjon om korrelasjonen mellom dei uavhengige variablane på førehand, vart likevel alle variablane teke med i analysen.

Modellane si utforming gjer at hellinga til regresjonslinja blir bestemt av Inntekt - den einaste kontinuerlege variabelen som er med. På grunn av dette (og alle dummy-variablane) vil regresjons-modellane bli "stive", dvs. ha vanskeleg for å fange opp variansen i data-materialet. Difor kan vi vente å få lave føynings-mål.

Dei estimerte koeffisientane vart testa ved hjelp av den vanlege t-observatoren. For å kunne bruke denne må SHI vere normalfordelt eller tilnærma normalfordelt. Som det går fram av avsnitt III.3 er ikkje SHI normalfordelt. På grunn av det store utvalet skulle det likevel gå bra å gjennomføre slike testar.

Effekten av ein binær uavhengig variabel blir målt i forhold til referansekategorien. For fritidsaktivitet er referansekategorien lett mosjon, slik at effekten av til dømes tung mosjon blir estimert som effekten av å drive tung mosjon samanlikna med å drive lett mosjon. Referansegruppa er sett saman av referansekategorien for kvar binær variabel. Samansetjinga av referansegruppa går fram av Tabell 1 og appendikset. Referansegruppa bør vere samansett av personar med gjennomsnittleg god helsetilstand. Valet av referansegruppa prøver å ta omsyn til dette, men det er vanskeleg å seie i kva grad denne målsetjinga er oppfylt.

Inndelinga i uavhengige og avhengige variablar som er gjort er problematisk pga. at eventuelle årsak/verknad forhold kan gå begge vegar (simultanitets-problem). Eit døme kan vere at analysen gjev ein positiv samanheng mellom helsetilstand og tung mosjon for menn som er 55 år eller eldre. Finn vi denne samanhengen fordi tung mosjon fører til god helse, eller fordi det berre er eldre menn med god helse som driv tung mosjon?

## 7. Kanonisk korrelasjonsanalyse

Ved hjelp av kanonisk korrelasjonsanalyse kan optimale vektorer på kvar av variablane i den samansette helsevariabelen SHI estimerast. Ein slik analyse kan vere ei naturleg vidareutvikling av den SHI med lik vekt på kvar av dei åtte del-variablane som vart brukt til regresjonane. I kanonisk korrelasjonsanalyse blir variablane inndelt i to grupper (t.d. avhengige variablar og uavhengige variablar). Analysen lager ein lineær kombinasjon av variabelane frå eine gruppa og ein lineær kombinasjon av variablane frå den andre gruppa slik at korrelasjonen mellom gruppene blir maksimert. Fleire slike kombinasjonar med ulike vektorer blir prøvd (sjå Nie et al. 1975, p.515-527).

Eit lite eksempel med to avhengige og to uavhengige variablar kan illustrere prinsippet:

$$(5) aY_{1i} + bY_{2i} = cX_{1i} + dX_{2i} \quad \text{for } i=1, \dots, n$$

der  $a$ ,  $b$ ,  $c$  og  $d$  er koeffisientane (vektene) som skal estimerast,  $Y_1$  og  $Y_2$  er avhengige variablar,  $X_1$  og  $X_2$  er uavhengige variablar og  $i$  er indeks for observasjon (person). I fyrste runde blir  $c$  og  $d$  estimerte ved minste kvadrat regresjon (som  $c$  og  $d$ ), medan  $a$  og  $b$  har verdien 1. Dei predikerte verdiane  $cX_{1i}$  og  $dX_{2i}$  blir så sett inn og høgre og venstre side av relasjonen bytt om:

$$(6) cX_{1i} + dX_{2i} = aY_{1i} + bY_{2i} \quad \text{for } i=1, \dots, n$$

No blir  $a$  og  $b$  estimert ved minste kvadrat regresjon og dei predikerte verdiane  $aY_{1i}$  og  $bY_{2i}$  sett inn. Deretter blir høgre og venstre side igjen bytt om. Dermed kan koeffisientane  $c$  og  $d$  estimerast på nytt. Estimata vil ha endra seg noko avdi  $a$  og  $b$  blei brukt i staden for  $a$  og  $b$ . Ved å gjenta denne prosedyren nokre gonger vil dei estimerte koeffisientane på begge sider stabilisere seg. Koeffisientane vil då vere optimale vektorer, dvs. dei vil maksimere korrelasjonen mellom  $(Y_1, Y_2)$  og  $(X_1, X_2)$ .

Fordi optimale vektorer (koeffisientar) i begge sett av variablar blir estimerte, vil ein større del av variansen i data-materialet bli forklart enn ved multippel regresjonsanalyse. Statistikk-pakken SPSS-X har ein underprosedyre for å utføre kanonisk korrelasjonsanalyse (underprosedyren 'DISCRIM' i 'MANOVA').

Vektene i SHI som blir estimerte er berre optimale ut frå datamaterialet som er brukt. For analyser på andre datamateriale vil dei berre vere optimale i den grad spørsmåla/variablane er dei same som i Helseundersøkelsen 1975 eller ligg nær opp til desse.

Kanonisk korrelasjonsanalyse vart berre utført på den additive modellen MODADD fordi logaritmen til SHI inngår i den multiplikative modellen MODMULT (logaritmen til ein sum kan ikkje løysast opp).

#### IV RESULTAT

##### 1. Forklaringskrafta til modellane

Forklaringskrafta til modellane er dårleg. I den samla analysen blir omlag 12% av variansen i SHI forklart av dei uavhengige variablane. For aldersgruppa 6-14 år blir berre 1-3.5% forklart, for aldersgruppa 15-54 år blir 10% forklart, og for aldersgruppa 55 år og eldre blir 11-16% forklart. I den yngste aldersgruppa ligg kvinner høgast, medan menn ligg høgast i den eldste aldersgruppa. Forklaringskraft ser ut til å stige med alderen. Den lave forklaringskrafta til modellane gjer resultatata usikre. Dei kan berre gje ein indikasjon på samvariasjon mellom SHI og dei uavhengige variablane. Forklaringskrafta til dei to modellane er nokså lik.

##### 2. Korrelasjon mellom dei uavhengige variablane

Ingen av korrelasjonane mellom dei uavhengige variablane er store. Delvis kjem dette av dei mange binære variablane som splitter opp dei opprinnelege variablane. Til dømes blir variabelen for region splitta opp i 11 binære variablar. Resultata tyder ikkje på alvorlege problem pga. korrelasjon mellom dei uavhengige variablane. Den største (positive) korrelasjonen finst mellom yrkesfeltet jord/skog/fiske og tungt kroppsarbeid (0.35 til 0.52). Elles ligg nokre av korrelasjonane over 0.2 i talverdi.

### 3. Hovudresultat frå analysen

På grunn av den lave forklaringskrafta til modellane blir berre forteiknet til dei signifikante koeffisientane presentert. Diskusjonen av resultatata tek utgangspunkt i forteiknet til desse signifikante koeffisientane. Resultata blir oppsummert i Tabell 1, der '+' viser at ein koeffisient er signifikant positiv og '-' viser at ein koeffisient er signifikant negativ. I tabellen er det markert om signifikansnivået er 5% eller 10%. Signifikansnivået viser til ein tosidig test av t-observatoren mot nullhypotesen: ingen effekt av den uavhengige variabelen. (Det er heile tida snakk om signifikante skilnader frå referansegruppa.) Sjølv om ein koeffisient er til dømes signifikant positiv, kan talverdien vere svært liten slik at ein kan sjå bort frå at den har noko effekt på SHI.

Fordi modellane gav nesten like resultat blir berre resultatata frå MODADD presenterte. Resultata frå MODADD er lettare å tolke og kan lettare samanliknast med resultatata frå den kanoniske korrelasjonsanalysen (som berre vart utført på MODADD).

For aldersgruppa 6-14 år er berre variablane 'Inntekt' og 'Region' med i regresjonane. På grunn av dette og fordi forklaringskrafta er spesielt lav for denne aldersgruppa, bør ein legge liten vekt på desse resultatata.



TABELL 1 SIGNIFIKANTE KOEFFISIENAR

	Alle	Menn	Menn	Menn	Kvinner	Kvinner	Kvinner
	6-14	15-54	55+	6-14	15-54	55+	
<b>FRITIDSAKTIVITET</b>							
Set i ro	-					-*	-
<u>Lett mosjon</u>							
Tung mosjon	+		+			+	
Trener hardt	**						
<b>YRKESFELT</b>							
<u>Teknisk/administrativt</u>							
Kontor/handel							
Jord/skog/fiske						**	
Transp./serv./militær						-	
Gruve/industri							
Husarb. m. innt.gj.arb.			-				
Husarb. u. innt.gj.arb.							
Elev/stud./vernepl.	+		**			+	
Pensjonist/anna	-		-			-	
Arbeidsufør	-						
<b>FYSISK AKT. I ARBEIDET</b>							
<u>Set i ro</u>							
Står og går mykje			**				**
Går og løfter mykje	-*						
Tungt kroppsarbeid							
<b>INNTEKT</b>							
			**			+	
<b>REGION</b>							
<u>Akershus</u>							
Oslo							
Aust- og Vest-Fold		**		**	**	**	
Hedmark/Oppland					-	+	
Buskerud/Telemark						+	
V.Agder/A.Agder/Rogal.	+	**				+	
Hordal./S.og Fjordane	+					+	
Møre og Romsdal	+		**	**	**	+	
Sør- og N.-Trøndelag						+	
Nordland	**					+	
Troms og Finnmark							
<b>RØYKEVANER</b>							
15+ sig./dag	-	-				-	
10-14 sig./dag	-					-	
1-9 sig./dag	-		-*			-	
Før dagl. røykar	-		-			-	-*
<u>Aldri røykt</u>							
<b>ALKOHOLFORBRUK</b>							
<u>Br. ikkje alkohol</u>							
Br. a. svært sjeldan	+					+	
Br. a. av og til	+					+	
Br. a. nesten dagleg			-				
<b>KORONAR HJARTESJUKD.</b>							
	-	-	-			-	

Signifikansnivå: Når ikkje merka av: 5%, \*: 10%  
Referansegruppa er merka ved understreking.

Forklaringskrafta til modellane er så lav at det er vanskeleg å trekke nokon konklusjon når det gjeld utgangshypotesen. Likevel gjev analysen ein indikasjon på støtte til hypotesen om at aukande fritidsaktivitet har ein positiv samanheng med helsetilstanden til folk. Derimot gjev den ingen indikasjon på at aukande inntekt har ein positiv samanheng med helsa til folk. For 'Yrkesfelt' og 'Fysisk aktivitet i arbeidet' er det signifikante utslag på helsetilstanden berre for nokre yrkesgrupper og typer fysisk aktivitet i arbeidet.

Det er 'Koronar hjartesyjukdom' og 'Yrkesfelt' (pensjonist) som slår sterkast ut av dei åtte uavhengige variablane. Den negative effekten av røyking er også signifikant. For menn mellom 15 og 54 år er det negative utslaget for dei som bruker alkohol nesten dagleg signifikant. Når det gjeld kvinner i same aldersgruppe er det mange signifikante utslag for region. For menn som er 55 år eller eldre er det eit signifikant (positivt) utslag på helsa av tung mosjon.

#### 4. Resultat frå den kanoniske korrelasjonsanalysen

Forklaringskrafta ved ein kanonisk korrelasjonsanalyse kan målast ved kvadratet på den kanoniske korrelasjonskoeffisient (sjå Nie et al. 1975, p.515-527). Denne kan samanliknast med kvadratet på den multiple korrelasjonskoeffisient frå ein regresjon.

Som venta gjekk forklaringskrafta opp for dei fleste gruppene. Den gjekk mest markert opp for kvinner som er 55 år eller eldre (frå 0.11 til 0.21), men er framleis svært lav. For heile utvalet og kvinner mellom 15 og 54 år gjekk forklaringskrafta ned. Dette ser ut til å komme av at ein del personar som var med under regresjonsanalysane ikkje kom med ved dei kanoniske korrelasjonsanalysane (personar med husarbeid utan inntektsgjevande arbeid).

Blant helse-variablane er det 'Fysisk evne ('førlighet')', som ser ut til å vere viktigast for å forklare variansen i dei uavhengige variablane. Andre viktige helse-variablar er: 'Nedsett Yrkesaktivitet pga sjukdom/skader', 'Fråver frå arbeidet siste 14 dagar', og 'Bruk av roande middel eller nervemedisinar'.

## 5. Oppsummering

Dei modellane som er brukt er i liten grad i stand til å fange opp variansen i datamaterialet. At dei fleste variablane inngår som binære variablar er ein viktig grunn til dette. Når ein bruker kategoriske data til regresjonsanalyser er dette eit av problema som dukkar opp. Som ein kunne vente på grunn av det store utvalet gjev regresjonane mange signifikante koeffisientar.

Analysen gjev ein indikasjon på støtte til hypotesen om at helsetilstanden til folk har ein positiv samanheng med aukande aktivitet (mosjon) i fritida. Denne hypotesen blir også støtta av enkle krysstabell-analyser som er gjennomført tidlegare. Desse tyder på at aktiv fritid i yngre alder har ein positiv effekt på helse (sjå kapittelet om friluftsliv i SS8: Miljø og Levekår, Rapport 1986/-). Analysen gjev ingen indikasjonar på at helsetilstanden stig med aukande inntekt. Når det gjeld yrkesfelt kjem studentar/elevar/vernepliktige ut med betre helsetilstand enn referansegruppa (teknisk/administrativt arbeid), medan pensjonistar kjem ut med dårlegare helsetilstand.

Tolkinga av slike resultat er vanskeleg avdi eventuelle årsak/verknads-forhold kan gå begge vegar. Er det ein positiv samanheng mellom helse og tung mosjon for menn som er 55 år eller eldre fordi tung mosjon fører til betre helse, eller fordi berre eldre menn med god helse driv med tung mosjon?

Resultata frå den kanoniske korrelasjonsanalysen tyder på at 'Fysisk evne ('førlighet')' er viktigast av helse-variablane; deretter 'Nedsett yrkesaktivitet pga sjukdom/skader', 'Fråver frå arbeidet s. 14 dagar' og 'Bruk av roande middel eller nervemedisinar'.

## V ERFARINGAR MED METODEN

Den lave forklaringskrafta til modellane viser at hypotese, data og metode ikkje går godt saman. Metoden ser ut til å passe best til kontinuerlege variablar.

For å teste hypotesen treng ein tidsseriar av individual-data. Det er eit dårleg utgangspunkt å bruke data frå berre eit tidspunkt når ein

eventuell effekt av til dømes fritidsaktivitet på helse tek tid og venteleg varierer med alderen. Dersom tidsseriar av individual-data blir tilgjengeleg, vil grunnlaget for å teste ein slik hypotese bli mykje betre. Panel-data vil vere eit steg på vegen. Ein vil få tilgang til panel-data dersom same eller delvis same utval blir brukt til nye data-innsamlingar, til dømes ved framtidige helsegranskingar.

Den kanskje viktigaste grunnen til at modellane gav lite er at det stort sett er nytta kategoriske data. Kor mykje det er mogeleg å få ut av ein statistisk analyse avheng av nivået på variablane. Det er lettare å gjennomføre ein liknande analyse med data på intervall-nivå eller forholdstal-nivå. Eit døme kan vere fysiske støy-data brukt i ein analyse av støy og helse.

Den samansette helseindikatoren SHI vart utvikla for denne analysen og er tilpassa metoden som er brukt. Spesielt for nokre aldersgrupper skiller SHI for lite mellom personar, slik at for mange blir liggande nær gjennomsnittet. SHI ser ikkje ut til å passe så godt til denne metoden, men ideen bør følgjast opp på andre område eller med andre data og metodar. Ved hjelp av ein samansett indikator kan fleire variablar som td. seier litt om helsetilstand, kombinerast til ein samansett indikator som gjev eit betre mål på helsetilstand. Brukt i kombinasjon med kanonisk korrelasjonsanalyse kan optimale vektorer i ein samansett indikator estimerast ut frå eit datamateriale. Ettersom valet av helse-variablar som inngår i SHI og vektene tilordna kvar kategori for nokre av variablane er gjort på subjektivt grunnlag, kan det vere verdt å prøve med andre kombinasjonar og variablar enn dei som er nytta i denne analysen. Dersom kontinuerlege variablar fanst, ville SHI ha passa betre til minste kvadrat regresjon. Ein ville i tillegg unngå å tilordne ei subjektiv vekt til kvar kategori av helse-variablane.

Ein annan grunn til den lave forklaringskrafta til modellane kan vere at viktige forklaringsvariablar ikkje er med i analysen. Likevel er nok dette ei mindre viktig forklaring. Ved framtidige analyser bør ein også legge vekt på å luke ut uavhengige variablar som er sterkt korrelerte.

Dei modellane som er brukt er så enkle som mogeleg for å bruke data til minste kvadrat regresjon. Den additive modellen MODADD og den multiplikative modellen MODMULT gav like resultat. Fordi MODADD er enklare å bruke og tolke, bør denne brukast til eventuelle nye

analyser. Det kan vere noko å vinne på å utvikle betre modellar.

Så lenge dei fleste data er kategoriske bør nok analysemetodar meir tilpassa kategoriske data ha prioritet framfor regresjonsanalyse. Til dømes analyse av fleirdimensjonale tabellar/log-lineær analyse.

For framtidige prosjekt på dette området er det to vegar å gå. Ein kan ta utgangspunkt i dei data som finst og prøve å bruke dei best mogeleg. I så fall må det leggast stor vekt på utvikling av metodar som kan utnytte kategoriske data, gjerne i kombinasjon med kontinuerlege variablar. Alternativet er å planlegge innhenting av data (datakvalitet) og metode ut frå den hypotesen som skal testast - altså ei integrert prosjektplanlegging.

## APPENDIKS

## VARIABLANE SOM INNGÅR I ANALYSEN

AVHENGIG VARIABEL er ein samansett helseindikator SHI. Denne er sett saman av 8 einskild-indikatorar, nemleg variablane:

\* V33 : Nedsett yrkesaktivitet pga av sjukdom eller likn. Svaret ja har fått verdien 0. Nei har fått verdien 1.

\* V136 : Fråver frå arbeidet siste 14 dagar pga sjukdom/skader. Målt i talet på dagar.

\* V195 : Kor lenge sidan siste kontakt med lækjar i eller utanfor institusjon. Svaralternativa er følgjande, der verdisetjinga er sett opp ved sida av:

. under 2 veker	: 0,1
. 2 til 4 veker	: 0,1
. 1 mnd til 0,5 år	: 1,0
. 0,5 år til 1 år	: 1,0
. 1 år til 5 år	: 10,0
. over 5 år	: 20,0
. aldri	: 100,0

Dei to fyrste klassane har same verdisetjing fordi vi reknar at forskjelen mellom dei er liten. Det same gjeld dei to neste klassane. Skaleringa tek delvis utgangspunkt i sprang i tid og delvis utgangspunkt i ei subjektiv vurdering av samanhengen mellom V195 og almann helsetilstand.

\* V311 : Fysisk evne ('førlighet'). Svaralternativa med verdisetjing er:

. gå i trapp, gå 15 min, springe 50 m	: 10
. gå i trapp, gå 15 min	: 3
. berre gå i trapper	: 2
. andre kombinasjonar med 1 fys. evne	: 1
. ingen fysisk evne	: 0

Klassifiseringa etter V311 er grov. Venteleg vil stort sett personar over 65 år komme i ein av klassane med nedsett fysisk evne. Difor har full fysisk evne fått verdien 10 med eit sprang ned til 3 for dei andre klassane.

\* S161 : Kontakt med helsetenesten. Variabelen er summen av:

. V161 : antal g. 1.besøk alm.pr.lækjar s.14 d.:kont.
. V162 : antal g. 1.besøk alm.pr.lækjar s.14 d.:heime
. V163 : antal g. gj.besøk alm.pr.lækj. s.14 d.:kont.
. V164 : antal g. gj.besøk alm.pr.lækj. s.14 d.:heime
. V165 : antal g. 1.besøk spesialist s.14 d.:kont.
. V166 : antal g. 1.besøk spesialist s.14 d.:heime
. V167 : antal g. gj.besøk spesialist s.14 d.:kont.
. V168 : antal g. gj.besøk spesialist s.14 d.:heime
. V169 : antal g. besøk lækj. på polyklinikk/sjukehus s.14 d.
. V174 : antal g. fysikalsk beh. utanf. inst. s.14 d.
. V175 : antal g. besøk kiropraktor siste 14 dagar

\* S170 : Kontakt med helsetenesten av meir alvorleg karakter.

Variabelen er summen av:

- . V170 :antal g. besøk av sjukepleiar s.14 dagar
- . V171 :antal g. innlagt i institusjon s.14 dagar
- . V182 :antal g. hos lækjar for observasjon s.14 d.

\* V227 : Har lækjar sagt at IO har hatt nervøse/psykiske plager.  
Ja har fått verdien 0 og nei har fått verdien 1.

\* V229 : Har IO brukt roande middel eller nervemedisin.  
Ja har fått verdien 0 og nei har fått verdien 1.

Den samansette helseindikatoren har formelen:

$$SHI = ZV33 - ZV136 + ZV195 + ZV311 - ZS161 - ZS170 + ZV227 + ZV229$$

Der prefikset Z viser at variabelen er standardisert ved å trekke frå det aritmetiske gjennomsnittet og dele på standardavviket. For kvar av alder/kjønns-gruppene er standardiseringa gjennomført ved å bruke gjennomsnittet og standardavviket innan gruppa. Desto høgare verdi dette målet på den almenne helsetilstanden til ein person har, desto betre er helsetilstanden.

Dei UAVHENGIGE VARIABLANE er:

\* I : Brutto hushaldsinntekt ifølgje skattelikninga pr forbrukseining i 100 kroner, 1975 kr. Denne er rekna ut ved å dele brutto hushaldsinntekt på summen av forbruksvektene til hushaldet (fyrste vaksne person: 1,0 ;fleire vaksne: kvar får vekta 0,7; barn under 16 år: kvart får vekta 0,5). Variabelen har namnet INNTEKT i databasen.

\* DF : Fritidsaktivitet siste 12 mnd. Denne er inndelt i 4 klassar. I modellen inngår variabelen gjennom 3 dummy-variablar:

- . DF<sub>1</sub> = 1 dersom set i ro  
= 0 elles
- . DF<sub>3</sub> = 1 dersom tung mosjon meir enn 4 timer i veka  
= 0 elles
- . DF<sub>4</sub> = 1 dersom trener hardt  
= 0 elles

Nullkategorien er dei som har lett mosjon meir enn 4 timar i veka. Variabelen har referansen V205 i databasen.

\* DB : Fysisk aktivitet i arbeidet. Denne er inndelt i fire klasser. Variabelen inngår i modellen gjennom følgjande dummy-variablar:

- . DB<sub>2</sub> = 1 dersom står og går mykje i arbeidet  
= 0 elles
- . DB<sub>3</sub> = 1 dersom går og løfter mykje i arbeidet  
= 0 elles
- . DB<sub>4</sub> = 1 dersom har tungt kroppsarbeid  
= 0 elles

I nullkategorien er personar som set i ro i arbeidet. Utgangsvariabelen har referansen V27 i databasen.

\* DS : Røykevaner og sigarettforbruk. Variabelen inngår i modellen gjennom følgjande dummy-variablar:

- . DS<sub>1</sub> = 1 dersom dagleg røykar, 15 el. fleire sig./dag  
= 0 elles
- . DS<sub>2</sub> = 1 dersom dagleg røykar, 10 - 14 sig./dag  
= 0 elles
- . DS<sub>3</sub> = 1 dersom dagleg røykar, 1 - 9 sig./dag  
= 0 elles
- . DS<sub>4</sub> = 1 dersom før dagleg røykar  
= 0 elles

Nullkategorien er samansett av dei som aldri har røykt. Variabelen har referansen ROYKING i databasen.

\* DA : Alkoholbruk. Variabelen inngår i modellen gjennom følgjande dummy-variablar:

- . DA<sub>2</sub> = 1 dersom bruker alkohol svært sjeldan  
= 0 elles
- . DA<sub>3</sub> = 1 dersom bruker alkohol av og til  
= 0 elles
- . DA<sub>4</sub> = 1 dersom bruker alkohol nesten dagleg  
= 0 elles

Nullkategorien er samansett av dei som ikkje bruker alkohol. Variabelen har referansen V271 i databasen.

\* DY : Yrkesfelt og type aktivitet. Variabelen inngår i modellen gjennom følgjande dummy-variablar:

- . DY<sub>2</sub> = 1 dersom kontor/handel  
= 0 elles
- . DY<sub>3</sub> = 1 dersom jord/skog/fiske  
= 0 elles
- . DY<sub>4</sub> = 1 dersom transport/service/militær  
= 0 elles
- . DY<sub>5</sub> = 1 dersom gruve/industri  
= 0 elles
- . DY<sub>6</sub> = 1 dersom husarbeid m. inntektsgj.arbeid  
= 0 elles
- . DY<sub>7</sub> = 1 dersom husarbeid u. inntektsgj.arbeid  
= 0 elles
- . DY<sub>8</sub> = 1 dersom elev/student/vernepliktig  
= 0 elles
- . DY<sub>9</sub> = 1 dersom pensjonist eller anna  
= 0 elles
- . DY<sub>10</sub> = 1 dersom arbeidsufør  
= 0 elles

Nullkategorien er teknisk/administrativt arbeid o.l. I databasen har variabelen referansen YRKE.

\* DK : Koronar hjartesyjukdom. Desse typene hjartesyjukdom er rekna med (ICE kode i parantes): akutt hjerteinfarkt/fersk (410), annan akutt/delvis akutt ischmetisk hjartesyjukdom (411), gammal hjerteinfarkt (412), koronar hjartesyjukdom utan at infarkt er nemt (413) og angina pectoris (414). Variabelen inngår som dummy-variabel:

- DK = 1 dersom har/har hatt ein av hjartesyjukdomane  
ovanfor
- = 0 dersom ein ikkje har/har hatt ein av desse  
syjukdomane

I databasen har variabelen namnet HJERTE.



\* DR : Region. Variabelen inngår i modellen gjennom følgende

dummy-variabler:

- . DR<sub>2</sub> = 1 dersom Oslo  
= 0 elles
- . DR<sub>3</sub> = 1 dersom Aust- eller Vestfold  
= 0 elles
- . DR<sub>4</sub> = 1 dersom Hedmark/Oppland  
= 0 elles
- . DR<sub>5</sub> = 1 dersom Buskerud/Telemark  
= 0 elles
- . DR<sub>6</sub> = 1 dersom Vest-Agder/Aust-Agder/Rogaland  
= 0 elles
- . DR<sub>7</sub> = 1 dersom Hordaland/Sogn og Fjordane  
= 0 elles
- . DR<sub>8</sub> = 1 dersom Møre og Romsdal  
= 0 elles
- . DR<sub>9</sub> = 1 dersom Sør- eller Nord-Trøndelag  
= 0 elles
- . DR<sub>10</sub> = 1 dersom Nordland  
= 0 elles
- . DR<sub>11</sub> = 1 dersom Troms eller Finnmark  
= 0 elles

Nullkategorien er Akershus. Variabelen har referansen V8 i databasen.

REFERANSAR

- Amundsen, H. Thoren (1976): Analysis of qualitative variables by means of binary regression.  
Memo 12.11.76 Institutt for Sosialøkonomi, Universitetet i Oslo
- Amundsen, H. Thoren (1984): Statistiske metoder for analyse av samvariasjon i kategoriske data  
SSB Rapport 1984/9
- Berg, R. L.(ed.) (1973): Health Status Indexes  
Proceedings of a Conference conducted by Health Services Research; Tucson, Arizona October 1 - 4, 1974  
Hospital Research and Educational Trust, Chicago 1973
- Elstad, J. I. (1983): Temaer omkring en generell sosial rapport  
INAS Rapport 1983:6, Oslo 1983
- Elstad, J. I. (1983): Levekår. Sammenheng og helhet  
INAS Rapport 1984:1, Oslo 1983
- Faye, A. og Herigstad, H. (1984): Friluftsliv i Norge 1970-82.  
SSB Rapport 1984/12
- Johnston, J. (1984): Econometric Methods  
International student edition, 3. ed. McGraw-Hill 1984
- Nie, N. H. et al. (1975): SPSS Statistical Package for the Social Sciences  
2nd ed., McGraw-Hill 1975
- Siring, E. og Spjøtvoll, E. (1984): Regresjonsanalyse med et stort antall variable  
SSB Rapport 1984/14
- SPSS Inc. (1983): User's Guide SPSS-X  
McGraw-Hill 1983
- Statistisk Sentralbyrå: Helseundersøkelsen 1968  
NOS A 321, Oslo 1969
- Statistisk Sentralbyrå: Friluftslivundersøkelse 1970  
NOS A 459, Oslo 1971
- Statistisk Sentralbyrå: Friluftslivundersøkelse 1974  
NOS A 725, Oslo 1975
- Statistisk Sentralbyrå: Levekår 1973  
NOS A 720, Oslo 1975
- Statistisk Sentralbyrå: Helseundersøkelse 1975  
NOS A 894, Oslo 1977
- Statistisk Sentralbyrå: Levekårsundersøkelsen 1980  
NOS B 320, Oslo 1982
- Statistisk Sentralbyrå: Levekårsundersøkelsen 1983  
NOS B 511, Oslo 1985
- Statistisk Sentralbyrå: Naturressurser og miljø 1984. Foreløpige nøkkeltall  
Rapport 1985/1

Statistisk Sentralbyrå: Miljø og Levekår  
Rapport 1986/- (under publisering)

Stewart, M. B. and Wallis, K. F. (1981): Introductory Econometrics  
2nd ed., Basil Blackwell, Oxford 1981