

Arbeidsnotater

S T A T I S T I S K S E N T R A L B Y R Å

10 69/3

Oslo, 3. mai 1969

INNTEKT OG UTDANNING

Inntektsvirkning av utdanning for hovedpersoner
i norske husholdninger 1967

av

Jan Isaksen

I n n h o l d

	Side
Virknninger av utdanning	1
Avgrensing av utdanningsbegrepet	1
Andre faktorer som kan være bestemmende for personlig inntekt	1
Inntektsforklaring i økonomisk teori	2
Mulige analysemetoder og funksjonsformer	4
Det statistiske materialet	6
Variabelliste	6
Kommentarer til de enkelte variable	7
Samspillvariable	10
Tolkning av modellen med samspill	12
Beregningsresultater og enkelte kommentarer	15
Litteraturliste	25
Vedlegg I (Utdrag av skjema for husholdnings- intervju. Boligundersøkelsen Oktober 1967)	26
Vedlegg II (Utdrag av kodeliste. Bolig- undersøkelsen Oktober 1967)	28

Dette arbeid er opprinnelig skrevet som spesialoppgave ved det sosialøkonomiske studium. Forfatteren har stått fritt i valg av opplegg og undersøkelsesmetoder. Arbeidet gjengis her med mindre endringer som forfatteren har ønsket å foreta. Synspunkter og konklusjoner står for forfatterens regning.

Ikke for offentliggjøring. Dette notat er et arbeidsdokument og kan siteres eller refereres bare etter spesiell tillatelse i hvert enkelt tilfelle. Synspunkter og konklusjoner kan ikke uten videre tas som uttrykk for Statistisk Sentralbyrås oppfatning.

VIRKNINGER AV UTDANNING

I denne undersøkelsen vil vi belyse utdanningens virkning på personlig inntekt. Utdanningens effekt på den enkeltes inntekt er selvfølgelig bare ett av de mange aspekter man kan trekke fram når man vil se på utdanningens totale innflytelse på samfunnet og på den enkelte. De produktivitetmessige fordeler et lands næringsliv kan ha av bedre utdannet arbeidskraft, blir f.eks. ofte påpekt.

Det er imidlertid en utbredt oppfatning at verdien av utdanning ligger mer på det kulturelle enn på det økonomiske plan. Dette vil si at de største virkninger av utdanning er ukvantifiserbare. Fra et økonomisk-teoretisk synspunkt blir forskjeller i utdanningsnivå mellom personer og mellom samfunn en stadig viktigere faktor for å forklare henholdsvis forskjeller i nasjonalinntekt og personlig inntekt. Enkelte hevder at man herved vurderer økonomiske virkninger av utdanning høyere enn andre virkninger. Det er imidlertid klart at selv om man finner utdanning viktig for å forklare personlig inntekt, innebærer dette ingen vurdering av hva som skal settes høyest av utdanningens effekter.

AVGRENSING AV UTDANNINGSBEGREPET

I videste forstand er utdanning alle impulser som en person mottar og som gjør ham bedre i stand til å forstå og gjennomskue ting. I denne undersøkelsen, hvor det økonomiske aspekt ved utdanningen er det sentrale, vil vi konsentrere oss mer om de trekk ved utdanningen som er yrkesrettede eller yrkesbetingende. Sett fra denne synsvinkel er det vanlig å dele utdanningen i skolemessig utdanning og etterutdanning som skjer på arbeidsplassen. Vi vil her ikke se på etterutdanning. Utdanning i vår terminologi blir da ensbetydende med almen-, universitets- og høyskoleutdanning.

ANDRE FAKTORER SOM KAN VÆRE BESTEMMENDE FOR DEN PERSONLIGE INNTEKT

Det virker ofte som om man ved "common sense" noenlunde sikkert kan fastslå en yrkesaktivs inntekt når man har greie på en del bestemte karakteristika ved personen. Man hører ofte uttalelser som: "Det skal litt til at en kvinne, selv med den utdanningen, skal tjene så mye", eller "han må vel tjene bra, han som har vært så lenge i samme jobb", eller "han som er tannlege må vel tjene godt". Her forutsetter man implisitt at slike forhold som kjønn, alder, yrke, sosial status ved siden av utdanning er bestemmende for en persons inntekt. Det er klart at skal man ved et statistisk materiale

belyse inntektsforskjeller på grunn av forskjeller i utdanning, så må man også trekke inn disse andre variable. De sistnevnte vil selvfølgelig også ha betydning for genereringen av materialet, og man må få skilt ut deres virkninger slik at utdanningens virkning kan sees mest mulig "rendyrket". Diskusjonen må derfor i stor grad dreie seg om flere inntektsbestemmende faktorer enn utdanningen.

INNTEKTSFORKLARING I ØKONOMISK TEORI

I økonomisk teori har man gjort forsøk på å forklare inntektsfordelingen i et samfunn. Dette har gått mest ut på å finne fram til fordelinger som passer til den inntektsfordelingskurve man kan observere. Bare i mindre grad har man pekt på faktorer som man mener forklarer denne fordeling. Når det gjelder spesielle studier over utdannings / inntektssammenhenger finner man for det meste forsøk på å fastlegge alderskurver for inntekten på ulike utdanningsnivåer. (Age-earnings profiles)

Personlig inntekt forklares ofte ut fra to alternative teorier. I den gjengse klassiske makromodell bestemmes arbeidslønnen av arbeidets grenseproduktivitet og en tilbudskurve for arbeidskraft under de vanlige klassiske forutsetninger. (Se (10) side 130). Hvis indifferensens lov gjelder i hele økonomien som betraktes, vil lønningene være like for alle arbeidstagere med samme arbeidsinnsats. Selvstendiges inntekt bestemmes på en vesenforskjellig måte.

I det vi vil kalle den sosiologiske modell, brukes en del karakteristika ved individet, såkalte sosio-økonomiske variable for å forklare den personlige inntekt. Her trekker man oftest ikke markedsforholdene eksplisitt inn.

En modell for inntektsbestemmelse som synes ganske plausibel kan være en blanding av de som er nevnt foran. Løst skissert kan man tenke seg økonomien oppdelt i mindre markeder, med en viss form for forbindelse mellom de enkelte. Lønnen på et bestemt marked bestemmes av tilbuds- og etterspørselsforholdene der. Hvilket arbeidsmarked den enkelte befinner seg på er imidlertid i stor grad bestemt av personlige egenskaper som f.eks. interesser og evner. Slike personlige egenskaper, ved siden av arv og tradisjon, vil også være utslagsgivende for valg av bransje, og dermed valg av fortjenestemuligheter for selvstendige. Til å forklare personlig inntekt kunne man etter dette tenke seg visse variabel som karakteriserte personlige egenskaper, og andre som karakteriserte forholdene på arbeidsmarkedet og/eller fortjenestemulighetene.

Variable som indikerer personlige egenskaper

Variable som gir et uttrykk for personlige egenskaper er vanskelige å få kvantifisert.

Kaller vi disse vanskelig kvantifiserbare for x_1, \dots, x_n kan vi skrive vår teori som: $y = f(x_1, \dots, x_n, W)$ hvor y står for personlig inntekt og W står for den eller de variable som skal karakterisere markedsforholdene på det enkelte marked.

Det vi så gjør er å gå ut fra at det finnes sosio-økonomiske variable z_1, \dots, z_m som er slik at de på en eller annen måte er funksjonelt bundne sammen med de opprinnelige variable, f.eks. på følgende måte:

$$\begin{aligned} x_1 &= g_1(z_1, \dots, z_m) \\ &\vdots \\ &\vdots \\ &\vdots \\ x_n &= g_n(z_1, \dots, z_m) \end{aligned}$$

Vi kan da skrive y som en annen funksjon av de kvantifiserbare z 'ene, slik:

$$y = f(g_1(z_1, \dots, z_m), \dots, g_n(z_1, \dots, z_m), W) = h(z_1, \dots, z_m, W)$$

Vi tror altså at de forhold som er med å bestemme hvilket marked den enkelte vil befinne seg i kan uttrykkes ved variable som står i et bestemt funksjonsforhold til de opprinnelige variable.

Markedsforholdene

De variable som skulle uttrykke markedsforholdene på vedkommende marked, måtte være indikatorer for tilbud og etterspørsel i markedet. Her vil det være naturlig å forutsette at det er en viss tidsforskyvning i tilpasningen på markedet, slik at markedsprisene først blir påvirket av tilbuds- og etterspørselsforholdene en viss tid etter at disse har forandret seg. Tverrsnittsdata vil således ikke være særlig egnet til å belyse denne sammenhengen.

Hvis den modell som her er skissert er riktig, må man kunne trekke den slutning at dersom forholdene på markedet er gitte, må den såkalte sosiologiske modell gi en noenlunde riktig beskrivelse av det som skjer. Spørsmålet er nå hvor realistisk det er å betrakte markedsforholdene og dermed lønns- og fortjenestemulighetene på det enkelte marked som gitte. Over et mindre tidsrom vil vel inntektene være temmelig konstante (i alle fall for arbeidstagere i tariffperiodene). En analyse ville fort bli uaktuell, skulle den bare ta for seg et slikt kort tidsrom. Imidlertid er det helst nivåhevninger som finner sted ved tariff-forandringer. (Dette er i alle fall ofte tilfellet i norsk

økonomi). Forholdet mellom de enkelte markeders nivåer må ansees å være mer konstant over tiden. Dette indikerer at en analyse som bygger på den sosiologiske modell alene, kanskje vil gi gale nivåer for inntekten absolutt sett, men at sammenhengene relativt sett vil bli noenlunde de samme over tiden.

Vi vil i det følgende behandle problemet ved hjelp av den sosiologiske modell. Analysen vil da selvfølgelig ha den begrensede gyldighet som ovenstående betraktninger gir uttrykk for.

I litteraturen finnes det endel forsøk på å forklare inntekten ut fra liknende modeller (se (2), (4), (6)). De fleste av disse holder de selvstendige inntektstakere utenfor analysen, idet man vel mener at eierinntekt genereres av en annen mekanisme. Inntekten forutsettes kanskje å skifte så sterkt med vekslende markedsforhold at en analyse av denne art har minimal interesse. I denne undersøkelsen er imidlertid de selvstendige inntektstakerne tatt med i analysen. De vil bare bli skilt ut ved at man antar at det er visse konstanter i funksjonsforholdet som får andre verdier for selvstendige enn for ikke-selvstendige. Et argument for et slikt opplegg er at man ikke tror at man har fått med alle forklarende faktorer, verken mhp. ansatte eller selvstendige, men kun de viktigste. Det er da vår tro at disse viktigste er de samme enten man er selvstendig eller ansatt.

MULIGE ANALYSEMETODER OG FUNKSJONSFORMER

Vårt utgangspunkt er altså at vi har diverse kvantitative eller klasseinndelbare variable som vi tror er av betydning for personlig inntekt. For å kunne sammenligne vår teori med data, må vi bestemme oss for en analysemetodikk.

Den mest brukte metode er tabelloppstillinger og gjennomsnittsberegninger. Det hefter imidlertid en del svakheter med slike metoder (se Artikler fra SSB nr. 6), særlig når antallet forklaringsvariable blir stort.

Det som antagelig ville være å foretrekke, var en mer eksplisitt matematisk formulert modell, hvor man kunne bruke de funksjonsformer som man vil tro gjør seg gjeldende.

Et hensyn man må ta når man har valgt en analysemetode av denne type er at den funksjonsform man velger, må være slik at man kan bruke statistiske metoder for å tallfeste koeffisienter.

Det er stort sett to forskjellige metoder som kan komme på tale. I regresjonsanalysen postulerer vi vanligvis den variable man vil forklare som en lineær funksjon av de forklaringsvariable. I variansanalysen opererer man vanligvis ikke med noen a priori antakelse om en funksjonssammenheng. Metodene

vil imidlertid falle sammen når man bruker binære variable for å uttrykke kvalitative variable, se (3). Det som nærmest har vært avgjørende for at det i denne analysen ble valgt en regresjonsmetode var at EDB-programmer for denne teknikk var best utviklet.

Vi skal i det følgende gå ut fra at inntektens sammenheng med sosioøkonomiske variable som utdanning, yrke, bosted etc. kan beskrives slik: Konstantleddet i en regresjonslikning med inntekt som venstresidevariabel antar en ny verdi når den sosioøkonomiske stilling for et individ endres. Et eksempel kan være når et individ går over fra gruppen "selvstendige" til gruppen "arbeidstakere". Dette får vi til ved å bruke en dummyteknikk som forklart i (3).

Den aldersvariable kunne vi også ha tatt inn i regresjonen på samme måte, dvs. delt den opp i aldersgrupper. Siden materialet ligger til rette for det, vil vi imidlertid bruke alderen som en kontinuerlig variabel.

Samspill mellom de variable

Ved å ta med såkalte samspillvariable i vår regresjonsmodell kan vi få undersøkt om utdanning har større inntektsvirkning i visse kombinasjoner med f.eks. yrke, alder eller kjønn enn sett isolert.

Vi vil f.eks. ha et uttrykk for samspill-virkningene mellom to egenskaper, som vi f.eks. kan kalle egenskap I og egenskap II. For hver av klassene som disse egenskapene er delt inn i (eks.: egenskapen utdanning er delt inn i 5 klasser) har vi en dummy-variabel som er 1 hvis individet befinner seg i nettop den klassen og 0 ellers. En variabel som gir et uttrykk for når et individ befinner seg både i en bestemt klasse i egenskap I og i en bestemt klasse i egenskap II, får vi ved å multiplisere de tilsvarende dummyer med hverandre. Den frambrakte variable kalles en "samspillvariabel" og koeffisienten foran den vil da gi uttrykk for hvor stort tillegg i konstantleddet det gir å tilhøre både den bestemte klasse i egenskap I og den bestemte klasse i egenskap II.

Det er klart at antall variable vil bli forvirrende stort hvis man dannet alle samspillvariable som det var mulighet for i modellen. Det gjelder her å begrense antallet av dem for å holde totalt antall variable i regresjonsberegningen innenfor en viss ramme. (Om reduksjon av antall samspillvariable se senere).

DET STATISTISKE MATERIALET

Materialet stammer fra en boligundersøkelse på intervjubasis utført ved Statistisk Sentralbyrås intervjukontor oktober 1967. Det ble tatt et stratifisert totrinnsutvalg på 3 000 norske husholdninger. Man kan lese nærmere om utvalgsmetodikken i "Boligundersøkelsen okt. 1967, Rapport fra kontoret for intervjuundersøkelser nr. 3, SSB 1968". Bortfallet var 185 og det var 32 observasjoner hvor man ikke fikk innhentet opplysning om inntekt. I boligundersøkelsen er husholdningen telleenheten, mens observasjonsenheten i denne undersøkelsen er hovedpersonen i husholdningen. (Det vil si det medlem av husholdningen som skaffer til veie den største del av husholdningens inntekt). Alle data refererer seg altså til hovedpersonen.

VARIABELLISTE¹⁾

Y	Inntekt
A	Hovedpersonens alder
A ²	Hovedpersonens alder kvadrert

- 1) Sammenliknet med boligundersøkelsens gruppering er vår gruppering grovere på enkelte punkter. I nedenstående tabell angir tallene i parentes hvilke grupper i boligundersøkelsens kodeliste (Vedlegg II) som er slått sammen. Tallet etter parentesen angir spørsmålets nummer på spørreskjemaet. (Vedlegg 1)

Referansegruppe for utdanning	(0,1) 3
Q ₂	(2,3) 3
Q ₃	(4,5) 3
Q ₄	(6,7) 3
Q ₅	(8) 3
R _y	(4,5,7) 1
Q ₆	(0,1,2,3) 1
Q ₇	(6,8,9) 1
Referansegruppe for bost. tetthet	(4) 9f
Q ₈	(1,2,3) 9f
Referansegruppe for uførhet	5a
Q ₉	(1-9) 5b
R _s	(2) 1
Q ₁₀	(1) 1
Q ₁₁	(3-9) 1
Referansegruppe for bolig u. oppv.	(3-7) 7
Q ₁₂	(1) 7
Q ₁₃	(2) 7

Kjønn

- Referansegruppe: Mann
 Q₁ Kvinne

Utdanning

- Referansegruppe: Bare 7-årig folkeskole (R_u)
 Q₂ Framhaldsskole
 Q₃ Middel- eller realskole
 Q₄ Eksamen artium
 Q₅ Universitet, høyskole

Yrkesgruppe

- Referansegruppe: Primærnæringer og industri (R_y)
 Q₆ Kontor- og handelsarbeide, tekn. vitenskaplig og adm. arbeide
 Q₇ Service, transport, kommunikasjon, militært og "uoppgitt"

Bostedstetthet

- Referansegruppe: Tettbygd strøk, strøk med over 20 000 innb.
 Q₈ Spredtbygd strøk og strøk med under 20 000 innbyggere

Uførhet

- Referansegruppe: Ikke ufør
 Q₉ Syk eller ufør

Sosial status

- Referansegruppe: Ansatt yrkesaktiv (R_s)
 Q₁₀ Selvstendig yrkesaktiv
 Q₁₁ Ikke yrkesaktiv

Bolig under oppveksten

- Referansegruppe: Bolig i hus sammen med andre familier, eller i rekkehus
 Q₁₂ Våningshus på gårdsbruk
 Q₁₃ Frittliggende enebolig

KOMMENTARER TIL DE ENKELTE VARIABLE

Inntekt

Angir antatt inntekt for hovedpersonen i følge statsskatteligningen 1966. (Dersom ikke oppført benyttes kommuneskatteligningen for samme år). Kodet i hele 100 kr. Inntekter over 99 000,- er kodet 990, men dette kan ikke regnes å ha noen særlig effekt på estimatene da ifølge inntektsstatistikk for 1965 bare ca. 1⁰⁰/0 av inntektstakere hadde høyere inntekt enn dette. En mer betydelig svakhet er at det ikke er mulig å skille ut de tilfelle hvor hovedpersonen er lignet sammen med en annen av familiemedlemmene. Felleslikning vil iflg. (6) medføre en inntekt 30 - 40 % høyere enn ellers. Det er

imidlertid lite trolig at dette skulle påvirke de innbyrdes størrelsesforhold på koeffisientene, i og med at fellesligning antakelig ikke vil forekomme så mye hyppigere sammen med noen egenskaper enn med andre.

De inntekter som i materialet er satt lik null, ialt 59 (dvs. ca. 2 % av utvalget), er holdt utenfor. Begrunnelsen for dette er følgende: I flere undersøkelser av inntektsfordelinger, se f.eks. (11), har man funnet at fordelingen som regel lar seg beskrive ganske godt ved en lognormal fordeling. Det kan se ut som om vi ikke kan bruke dette synspunkt på det foreliggende materiale da vi har null-inntekter her og logaritmen til null ikke er definert. Mot dette kan man innvende at inntekten bare tilsynelatende er 0. Man har imidlertid ikke muligheter for å korrigere for dette ved å legge til den inntekt som var for lav til å bli iliknet skatt hos den enkelte. Det man eventuelt da kunne gjøre var f.eks. å legge til et konstant beløp til hver inntekt (se (11) kap. 9). Dette konstante beløp ville i så fall bli helt vilkårlig og estimatene ville variere ettersom beløpet ble forandret. Vi velger å kutte ut observasjonene i stedet, dette medfører i alle fall at man kan ha en begrunnet mening om hvilken vei feilen går.

Det er trolig at de som ikke har skattbar inntekt ikke er familieforsørgere, og dette medfører stor sannsynlighet for at de er enten i de helt unge eller i de meget eldre aldersklasser. Dette igjen skulle indikere at de befant seg på et relativt lavt utdanningstrinn, og når vi da holder dem utenfor utvalget vil dette medføre for høye estimater for de laveste utdanningsgruppene.

Hovedpersonens alder

Denne anses i alle tidligere undersøkelser av liknende art å ha stor betydning for størrelsen av den personlige inntekt. Se f.eks. (4), (6) og (9). Den svakhet som kanskje finnes i dette materialet er at det kan tenkes at særlig yngre folk i arbeidslivet som egentlig er selvforsørgende men som ikke har egen husholdning bare kommer inn i undersøkelsen som husholdningsmedlem og ikke som selvstendig inntektstaker.

Hovedpersonens alder kvadrert

Begrunnelsen for å ta denne variable med er at inntektens forløp med alderen ofte anses å ha form som et annengradspolynom. (Dvs. $Y = a + bA + cA^2$ hvor man som regel regner med at c er negativ).

Kjønn

Kjønnsforskjeller ansees å ha virkning på personlig inntekt.

Utdanning

Materialet gav mulighet for å splitte opp slik at man kunne få skilt ut spesialutdanning. Ønsket om å holde antall variable innen en ramme som var overkommelig for regresjonsprogrammet, og som gjorde det lett å holde oversikten over beregningsresultatene, førte til ovenstående gruppering.

Yrkesgruppe

Det må antas å ha sterk virkning på inntektens størrelse hvilken yrkesgruppe en person tilhører.

Referansegruppe for yrke: Primærnæring og industri: omfatter industri, bygg og anlegg, jordbruk, skogbruk, fiske, gruver og sprengningsvirksomhet.

Bostedstetthet

Det har vært brukt å dele materialet inn i geografiske områder, da man antar at det er inntektsforskjeller mellom områdene. Det er mulig at disse forskjellene mer skyldes at industrialiseringen og dermed urbaniseringen er forskjellig i de forskjellige landsdelene. Vi har derfor skiftet ut den geografiske variable med et mer direkte uttrykk for samme faktor.

Uførhet

Hovedpersonens grad av arbeidsførhet er det nødvendig å ta hensyn til, da det er klart at en ufør person gjennomgående vil ha mindre inntekt enn en som er fullt arbeidsfør.

Sosial status

Her har spesielt tredelingen: ansatt, selvstendig, ikke yrkesaktiv, interesse.

Bolig under oppveksten

Denne variable er tatt med ut fra tanken om at den sosiale bakgrunn spiller en rolle for en persons inntekt. (Jfr. eplet og stammen, som vel også har en viss relevans inntektsmessig sett). Det som er vanskelig å si er i hvor sterk grad kjennetegnet bolig kan sies å karakterisere "sosial bakgrunn". Noen bedre indikasjon for denne faktor fantes desverre ikke i datagrunnlaget.

SAMSPILLVARIABLE

Man kunne prinsipielt danne samspillvariable mellom alle de egenskaper (alder, utdanning etc.) som vi har tatt med i vår modell. Et slikt opplegg forbyr imidlertid seg selv ved det voldsomt store antall nye variable en måtte danne. Vi har derfor plukket ut en del variable som vi a priori synes det er interessant å se på samspillet mellom. Disse variable er:

Alder/utdanning

Sosial status/utdanning

Yrke/utdanning

Alder/yrke

En ytterligere reduksjon er kommet i stand ved at vi har sett bort fra variable som er av liten interesse etter et a priori resonnement, og ved sammenslagning som det fremgår av nedenstående tabeller.

x = aktuell kombinasjon

o = Kombinasjonen ansees som lite interessant

 = sammenslagning av klassene

Yrke/utdanning

	R _u	Q ₂	Q ₃	Q ₄	Q ₅
R _y	x	x	x	o	o
Q ₈	o	<u>x</u>	<u>x</u>	x	o
Q ₇	o	o	<u>x</u>	<u>x</u>	x

Sosial status/utdanning

	R _u	Q ₂	Q ₃	Q ₄	Q ₅
R _s	<u>x</u>	<u>x</u>	<u>x</u>	x	x
Q ₁₁	<u>x</u>	<u>x</u>	<u>x</u>	x	x
Q ₁₂	o	o	o	o	o

Kommentarer til reduksjonene

Q_{12} dvs. gruppen "ikke yrkesaktiv" kan vel ikke være særlig interessant å betrakte sammen med utdanning når det gjelder den virkning en slik sammensatt faktor får på inntekten.

Kombinasjonene i øverste høyre og nederste venstre hjørne av tabellen "Yrke/utdannelse" bringes ikke inn, fordi det ikke virker interessant å se på yrkes/utdanningskombinasjoner som antakelig svært sjelden forekommer i praksis.

Sammenslagningene (markert ved klammer) er foretatt for å holde antall variable nede. Det er klart at slike aggregeringer alltid kan diskuteres, men så lenge det gjelder å påvise eventuelt om virkningen av samspill kan sies å være vesentlig, vil det ikke være av største viktighet hvordan klassene er slått sammen.

Definisjon av de samspillvariable

Kombinasjon av:	Variabelsymbol	Dannes av produktet av gruppevariablene:
Alder/yrke	S_1	R_y og A
	S_2	Q_7 og A
	S_3	Q_8 og A
Alder/utdanning	S_4	R_u og A
	S_5	Q_2 og A
	S_6	Q_3 og A
	S_7	Q_4 og A
	S_8	Q_5 og A
Yrke/utdanning	S_9	R_u og R_y
	S_{10}	Q_2 og R_y
	S_{11}	Q_3 og R_y
	S_{12}	Q_2 og Q_8

(forts.)

Kombinasjon av:(forts.)	Variabel-symbol	Dannes av produktet av gruppevariablene:
Yrke/utdanning (forts.)	S ₁₃	Q ₃ + Q ₄ og Q ₈
	S ₁₄	Q ₃ + Q ₄ og Q ₇
	S ₁₅	Q ₅ og Q ₇
Sosial status/utdanning	S ₁₆	R _u + Q ₂ + Q ₃ og R _s
	S ₁₇	Q ₄ og R _s
	S ₁₈	Q ₅ og R _s
	S ₁₉	R _u + Q ₂ + Q ₃ og Q ₁₁
	S ₂₀	Q ₄ og Q ₁₁
	S ₂₁	Q ₅ og Q ₁₁

TOLKNING AV MODELLEN MED SAMSPILL

I det følgende vil vi betrakte en modell for inntektsforklaring som analytisk er formulert slik:

$$(1) \quad Y = a_1 + a_2 A + a_3 A^2 + \sum_{i=1}^{13} b_i Q_i + \sum_{j=1}^{21} c_j S_j$$

Ved å innføre samspillvariable på den måten det er gjort her, og altså utelukke samspillet mellom enkelte grupper, vil man tro at man kan få riktigere tilpassing av et regresjonshyperplan til materialet.

Imidlertid må man bruke en annen tolkning såvel av konstantleddet a_1 som av de enkelte koeffisienter når den tilhørende variable også er tatt med i en samspillkombinasjon.

Vi kan ta utgangspunkt i modellen uten samspillvariable. (Dvs. de siste 21 ledd i (1) sløyfes). Konstantleddet a_1 må da tolkes som den inntekt en person vil få dersom han tilhører alle referansegruppene. Vi kan skrive konstantleddet som:

$$(2) \quad a_1 = d_{u u} R_u + d_{y y} R_y + d_{b b} R_b + d_{h h} R_h + d_{s s} R_s + d_{o o} R_o + d_{k k} R_k$$

d 'ene er konstanter. R 'ene kan vi betrakte som dummy-variable som gir uttrykk for at individet befinner seg i en eller annen referansegruppe.

(Fotskriftene viser hvilken referansegruppe den variable representerer.)¹⁾

Det vi gjør i regresjonsopplegget uten samspill, er å sette alle $R \equiv 1$ slik at $a_1 = d_u + d_y + d_b + d_h + d_s + d_o + d_k$

I modellen med samspill (hele (1)), lar vi derimot enkelte av R 'ene fungere som vanlige dummies og multipliserer dem med andre variable slik at vi frembringer samspillvariable.

I vår modell ser vi på samspillet mellom gruppene utdanning, yrke og sosial status. De variable vi ikke setter $\equiv 1$ er da R_u , R_y og R_s . For modellen gjengitt i (1) vil da uttrykket $d_u R_u + d_y R_y + d_s R_s + d_b + d_h + d_o + d_k$ hvor R_u , R_y , R_s oppfattes som vanlige dummies, bli det uttrykk som blir sammenliknbart med a_1 i modellen uten samspill. Vi får på en måte bare leddene d_b , d_h , d_o og d_k tilbake i konstantleddet.

Ved at vi danner samspillvariable av R_u , R_y og R_s blir imidlertid ikke d_u , d_y og d_s konstante, men de vil variere alt etter hvilken konstellasjon av egenskaper den enkelte person har. Ved å bruke definisjonene av de samspillvariable kunne vi få et uttrykk for de sammensatte "koeffisienter" d_u , d_y og d_s hvis vi samlet alle de ledd i regresjonen som hadde hhv. R_u , R_y og R_s som en komponent. For d_u f.eks. er disse leddene: $c_4 R_u A$, $c_9 R_u R_y$, $c_{16} (R_u + Q_2 + Q_3) R_s$, $c_{19} (R_u + Q_2 + Q_3) Q_{10}$

Av dette ser vi pga. aggregeringen som ligger i de to siste uttrykkene går det ikke an å legge sammen disse uttrykkene og sette R_u utenfor en parentes. Man får derfor ikke et uttrykk som er direkte sammenliknbart med d_u . Tilsvarende gjelder for d_s . For d_y , derimot kan vi finne et slikt uttrykk:

$R_y (c_1 A + c_9 R_u + c_{10} Q_2 + c_{11} Q_3)$ hvor parentesens etter R_y er sammenliknbar med d_y .

Vi ser herav at et uttrykk for konstanten i opplegget med samspill som skal kunne sammenliknes med a_1 i opplegget uten samspill vil variere med de andre variable hvis man i det hele tatt aritmetrisk kan komme fram til et slikt uttrykk. Uten å vite noe om størrelsesordenen på d_u , d_y og d_s må man i et hvert fall tro at konstanten i samspillopplegget blir mindre (i tallverdi enn den i et vanlig

- 1) u = utdanning
 y = yrke
 b = bostedsforhold
 h = handikap/ikke handikap
 s = sosial status
 o = bolig under oppveksten
 k = kjønn

opplegg.

Tolkningen av de andre koeffisientene i opplegget med samspill vil også bli problematisk, og vil variere fra den ene konstellasjon av egenskaper til den andre. Vi kan som et eksempel se på leddet a_2A ("aldersleddet"). I samspillopplegget vil det som tilsvarende dette i det enkle opplegget bli:

$$A(a_2 + c_{1R}R_y + c_{2Q}Q_6 + c_{3Q}Q_7 + c_{4R}R_u + c_{5Q}Q_2 + c_{6Q}Q_3 + c_{7Q}Q_4 + c_{8Q}Q_5)$$

Inntektens variasjon med alderen vil her bli avhengig av utdanning og yrke. Ved å derivere regresjonslikningen mhp A får vi:

$$\frac{\partial Y}{\partial A} = a_2 + c_{1R}R_y + c_{2Q}Q_6 + c_{3Q}Q_7 + c_{4R}R_u + c_{5Q}Q_2 + c_{6Q}Q_3 + c_{7Q}Q_4 + c_{8Q}Q_5 + 2Aa_3$$

den annenderiverte blir:

$$\frac{\partial^2 Y}{\partial A^2} = 2a_3$$

Vår modell innebærer altså at stigningen i inntekt med alder er avhengig av utdanning og yrke. Derimot har disse faktorer ingen innflytelse på aksellerasjonen av inntekten.

Modellens stokastiske forutsetninger

Den teoretiske fordeling

De undersøkelser som har vært gjort tidligere på dette området, tyder på at inntektsfordelingen er skjev. Altså kan man ikke forutsette at man har med en normalfordeling å gjøre, hvilket ville medføre at vi kunne trekke endel sterkere konklusjoner ut fra våre estimater.

Det er mye som tyder på at en lognormal fordeling er en god tilnærming til den inntektsfordeling man ofte finner i et samfunn. (Se f.eks. (11)). Vi har her funnet det plausibelt å anta at inntektsfordelingen kan ha denne form. Hvis man så tar logaritmen til inntekten vil denne være normalfordelt og medføre endel gode egenskaper for våre estimater. Å ta logaritmen vil også si at de inntektsbestemmende faktorer øker eller reduserer inntekten med en multiplikativ faktor. Etter dette vil den modell vi anvender se slik ut i matematisk formulering:

$$\ln Y_r = a_1 + A_r^1 a_2 + a_3 A_r^2 + \sum_{i=1}^{13} b_i Q_{ir} + \sum_{j=1}^{21} c_j S_{jr} + U_r$$

a_1, a_2, a_3, b_i, c_i er konstante koeffisienter som vi vil estimere. Fotskrift r står for individ nr. r .

Vi synes nå at det er plausibelt å gjøre følgende forutsetninger om restleddene U_r

- 1) $EU_r = 0$ og $\text{var } U_r = \sigma^2$ for alle r
- 2) U_r 'ene er stokastisk uavhengige

Vi kan nå bruke minste kvadraters metode til å estimere de konstante koeffisienter i vår modell. (Denne metode viser seg å falle sammen med sannsynlighetsmaksimeringsmetoden under våre forutsetninger).

Problemet med interkorrelasjon i de forklaringsvariable

I (4) nevnes problemet med interkorrelasjon i de forklaringsvariable. Det er lett i teorien å tenke seg at sosio-økonomiske variable kan henge sammen i en slags kausalkjede. F.eks. kan den utdanning en person får være bestemt på grunnlag av foreldrenes sosiale status og vedkommendes kjønn. Utdanningen kan igjen være bestemmende for hvilket yrke og hvilken næringsgren vedkommende skal arbeide i. Det kan tenkes som et siste ledd i kjeden, at yrke og næringsgren kan virke retningsgivende for personens plassering i en sosial-status-gruppering.

Hvilket omfang interkorrelasjonsproblemet har kan man lese av i korrelasjonsmatrisen mellom de forklaringsvariable. Det er imidlertid klart at med utstrakt bruk av klassifikasjonsvariable vil tallene i korrelasjonsmatrisen være svært avhengige av hvordan man grupperer data. Ved å nytte alternative klassifiseringsmetoder vil man kunne få helt forskjellige resultater mht. korrelasjonsmatrisen.

For å gjøre dette problemet mindre kan man selvfølgelig unnlate å ta med karakteristika som det anses å være en høy grad av sammenheng mellom. I denne undersøkelsen vil vi ikke forsøke en slik fremgangsmåte, men har prøvet å velge vår klasseinndeling slik at problemet blir minst mulig. Gruppeinndelingen behøver ikke nødvendigvis bli uinteressant selv om den er ordnet slik at interkorrelasjonen blir liten. Om interkorrelasjonsproblemet allikevel er av alvorlig art, vil dette merkes på kvaliteten av våre estimater.

BEREGNINGSRISULTATER OG ENKELTE KOMMENTARER

Beregningene er foretatt på Statistisk Sentralbyrås EDB-anlegg, ved hjelp av et standardprogram for lineær regresjon. (Se (7)).

Nedenstående tabell viser det empiriske gjennomsnitt og standardavvik for alle de variable.

Variabel	Gjennomsnitt	Standardavvik
Y	9,72300	
A	51,611	15,428
A ²	2901,723	1628,715
Q ₁	0,146	0,353
Q ₂	0,177	0,382
Q ₃	0,093	0,290
Q ₄	0,050	0,218
Q ₅	0,030	0,172
Q ₆	0,195	0,397
Q ₇	0,152	0,359
Q ₈	0,870	0,337
Q ₉	0,103	0,304
Q ₁₀	0,213	0,409
Q ₁₁	0,189	0,392
Q ₁₂	0,489	0,500
Q ₁₃	0,257	0,437
S ₁	22,460	25,720
S ₂	8,904	18,967
S ₃	6,929	17,076
S ₄	35,751	28,633
S ₅	7,879	18,299
S ₆	4,425	14,770
S ₇	2,195	10,027
S ₈	1,362	7,967
S ₉	0,350	0,477
S ₁₀	0,095	0,293
S ₁₁	0,015	0,120
S ₁₂	0,028	0,166
S ₁₃	0,030	0,172
S ₁₄	0,075	0,263
S ₁₅	0,028	0,166
S ₁₆	0,535	0,499
S ₁₇	0,039	0,193
S ₁₈	0,024	0,154
S ₁₉	0,199	0,399
S ₂₀	0,006	0,074
S ₂₁	0,006	0,074

Bortsett fra beregningen av koeffisientene i vår hovedmodell hvor samspilleffektene er forsøkt trukket inn, har vi beregnet koeffisientene i en tilsvarende modell uten samspill. En sammenligning av disse to modeller viser en del interessante trekk.

Testen det refereres til i resultat-tabellen nedenfor, er en test av hypotesen: H_0 : Koeffisienten = 0 mot H_1 : koeffisienten \neq 0. De tilfelle da null-hypotesen kunne forkastes ved et signifikansnivå på 5% er markert med S i tabellen. (Se (7) side 13, pkt. 11). Koeffisientene er også regnet om til forholdstall, disse angir (multiplisert med 100) i prosent vedkommende gruppes inntekt i forhold til referansegruppens.

Koeffisient	Estimat angitt i logaritmer	Estimat (Forholdstall)	Signifikant/ ikke signifikant
a_1	8,78091	6507,15	
a_2	0,02862		
a_3	-0,00031		S
b_1	-0,38784	0,679	S
b_2	0,08578	1,090	
b_3	-0,10514	0,900	
b_4	-0,77178	0,462	S
b_5	-0,39874	0,671	
b_6	-0,08040	0,932	
b_7	0,08145	1,084	
b_8	-0,11546	0,891	S
b_9	-0,19469	0,823	S
b_{10}	1,92038	6,823	
b_{11}	-0,30217	0,739	
b_{12}	-0,07921	0,924	S
b_{13}	-0,03185	0,969	
c_1	-0,01093	0,989	S
c_2	-0,00728	0,993	
c_3	-0,01337	0,987	S
c_4	0,00661	1,007	
c_5	0,00939	1,009	
c_6	0,01646	1,017	
c_7	0,02513	1,026	
c_8	0,02517	1,026	
c_9	-0,11815	0,889	
c_{10}	-0,23613	0,790	

(forts.)

Koeffisient	Estimat angitt i logaritmer	Estimat (Forholdstall)	Signifikant/ ikke signifikant (forts.)
c ₁₁	-0,15832	0,854	
c ₁₂	-0,23938	0,787	
c ₁₃	-0,17306	0,841	
c ₁₄	-0,09935	0,905	
c ₁₅	-0,29818	0,742	
c ₁₆	0,88502	2,561	
c ₁₇	1,24509	3,473	
c ₁₈	1,28929	3,631	
c ₁₉	-1,22657	0,239	
c ₂₀	-0,88835	0,411	
c ₂₁	-0,58322	0,558	

Modell I

Den multiple korrelasjonskoeffisient ble i modellen med 36 forklaringsvariable 0,57809. Dette kan tyde på at litt over halvparten av variasjonen i materialet er forklart ved vår modell. Man skal imidlertid være forsiktig med å komme med tolkninger av den multiple korrelasjonskoeffisient når modellen er bygget på såpass løse betraktninger som vår. Ut fra den antakelse at den multiple korrelasjonskoeffisient i alle fall sier noe om hvorvidt det går an å forklare inntekt ved hjelp av sosio-økonomiske variable, kunne det være interessant å sammenlikne med liknende undersøkelser som er foretatt.

Undersøkelse	R ²	Antall forklaringsvariable	Antall observasjoner
Østby (9)	0,6492	25	24.142
Vatle (6)	0,1700	18	9.035
Modell I	0,3358	36	2.724
Modell II	0,3144	14	2.724

På grunnlag av tabellen ovenfor må man kunne si at de resultatene en her er kommet til mht. multiplere korrelasjon, hverken er positivt eller negativt overraskende.

Det som vel er mest iøynefallende ved resultatene, er de relativt få koeffisientene i resultat-tabellen som er signifikant forskjellige fra 0. Dette kan dels være en følge av å bruke en for sterk klasseoppdeling, fordi man da får for få observasjoner i hver klasse til å få pålitelige estimater. (Se (12) side 227).

Et ganske overraskende resultat er at alder ikke skulle ha signifikant virkning på inntekten. Derimot har man for estimatene på c_1 , c_2 og c_3 fått et signifikant resultat som kan tolkes dithen at man må se alderens virkning på inntekt i sammenheng med hvilket yrke vedkommende har. At disse signifikante resultater går i en annen retning enn man kunne tenke seg, er lite oppmuntrende.

For alder kvadrert er estimatet på koeffisienten det man kunne vente seg ifølge teoretiske betraktninger og tidligere beregninger. Likeledes kan man si at resultatet for b_1 , dvs. det fradrag i inntekt som gis for å være kvinne, støtter opp om teorien om at vi står overfor forskjeller i inntektsnivå mellom kjønnene, selv etter at vi har tatt hensyn til andre forskjeller. Det må tilføyes at en del av denne forskjell her kan skyldes at kvinnen mer hyppig tar deltidsarbeid, en faktor som vårt materiale dessverre ikke registrerer. Begge de ovennevnte resultater er signifikante.

Det ene signifikante resultat man fikk for utdanningsgrupperingen må vel også karakteriseres som "rart". Det ser ut som at det å ha eksamen artium medfører en nedgang i inntekt sammenliknet med referansegruppen som er "folkeskole". Man skulle vente at alle utdanningsgruppene over folkeskolen ville få et tillegg i sin inntekt. I nedenstående tabell er alle de (usignifikante) samspillkoeffisientene som har med utdanning å gjøre trukket inn i et eksempel for å belyse utdanningens virkning hos en person som tilhører referansegruppene overalt unntatt i utdanning.

Framhaldsskole	$Q_2 b_2$ 0,08578	$c_2 Q_2 R_y$ -0,23613	$c_{16}((R_u + Q_2 + Q_3)R_s)$ 0,88502	Total virkning av utd. 0,73467
Middelskole	$Q_3 b_3$ -0,10514	$c_{11} Q_3 R_y$ -0,15832	$c_{16}((R_u + Q_2 + Q_3)R_s)$ 0,88502	0,62156
Eksamen artium	$Q_4 b_4$ -0,77178		$c_{17}(Q_4 R_s)$ 1,24509	0,47331

(forts.)

(forts.)			Total virkning av utd.
Univ. og høgsk.	Q_5^b -0,39874	$c_{18}(Q_5^R)$ 1,28929	0,89055

Man får nå et tillegg til inntekten for utdanning ut over folkeskole-nivå, hvilket er rimelig. Det uventede er nå rekkefølgen av utdanningstypene mhp. hvilken som gir størst inntekt. Mens man skulle vente rekkefølgen: universitet og høgskole (størst), gymnas (nest størst), realskole og framhalds-skole, får man: Universitet og høgskole, framhaldskole, realskole og gymnas. En grunn til dette som kan tenkes er at den alderssammensetning vårt materiale har, antagelig med stor vekt på høyere aldersklasser er det mulig at en person selv bare med framhaldsskole e.l. nå kan tjene svært godt. Man kan si at hans utdanning målt med hans tids utdanningsmålestokk er ganske bra, og det er dette han blir belønnet for.

Estimatene på b_8 og b_9 er begge signifikante, og resultatene var de ventede. Ad b_8 : folk som bor i bymessige strøk, dvs. med over 20 000 innbyggere, har ca. 10 - 15% større inntekt enn dem som bor i spredt bebygde strøk og strøk med under 20 000 innbyggere. Ad b_9 : personer med handikap har bare 80 - 85% av den inntekt som de uten handikap har.

Oppdelingen i klasser for sosial status synes ikke å gi noe større bidrag til forståelse av variasjon i inntekten, da hverken estimatet av b_{10} eller b_{11} er signifikant forskjellig fra 0.

Et av de forholdsvis få signifikante resultater er også at det ser ut til å ha en viss betydning i hva slag bolig inntektstakeren er vokst opp. I forhold til referansegruppen gir oppvekst på "våningshus og gårdsbruk" et fra- drag i inntekten på ca. 8%. Det kan tyde på at vår teori om en sammenheng mellom inntekt og foreldres status har noe for seg. Forskjellene er imidlertid ikke så store at oppdagelsen kan sies å være særlig overraskende. Det kan tenkes at en hvilken som helst variabel kunne gi like mye "forklaring".

Når unntas c_1 og c_3 var ikke ett eneste av estimatene for samspill- koeffisientene signifikant forskjellig fra 0. Man må vel herav kunne slutte at denne type variable ikke gir noe vesentlig bidrag til forklaring av inn- tekten. Dette kan tyde på at enten er denne bestemte måte å innføre sam- spillet i modellen på ikke hensiktsmessig, og bør avstås fra i senere under- søkelser, eller at et slikt samspill som man forestilte seg bare i svært liten grad er tilstede. Også hvis det siste er tilfelle bør utfallet av denne undersøkelsen medføre at man i senere undersøkelser av samme type heller beskjeftiger seg med de direkte variable. Det skulle altså ikke medføre alt

for stor forenkling å se bort fra samspillet. Dette resultatet finner vi også referert i (4).

Modell II (uten samspill)

Koeffisient	Resultattabell		Signifikant/ ikke signifikant
	Estimat angitt i logaritmer	Estimat (Forholdstall)	
a_2	0,02899		S
a_3	-0,00030		S
b_1	-0,39998	0,671	S
b_2	0,15575	1,169	S
b_3	0,30549	1,357	S
b_4	0,31332	1,368	S
b_5	0,61280	1,846	S
b_6	0,21221	1,236	S
b_7	0,05325	1,055	
b_8	-0,11724	0,889	S
b_9	-0,22645	0,797	S
b_{10}	-0,19953	0,819	S
b_{11}	-0,36368	0,695	S
b_{12}	-0,08857	0,915	S
b_{13}	-0,03930	0,961	
a_1	9,29577	10890,54	

I denne modellen finner vi omtrent alle estimatene signifikant forskjellige fra 0, og en multippel korrelasjonskoeffisient på 0,56075. Det siste kan vi ta som enda et indisium på hvor lite forklaringskraft som ligger i de 21 samspillvariable. Resultatene ligger også i større grad enn i hovedmodellen nær opp til det man kunne vente seg i følge "common sense".

Aldersvariasjonen er signifikant for begge de aldersvariables vedkommende, og avviker størrelsesmessig nesten ikke fra koeffisientene i hovedmodellen. Fradraget i inntekten for kvinner er også temmlig nøyaktig like stort her som i hovedmodellen.

Forskjeller i inntekt pga. forskjellig utdanning viser seg her atskillig mer som ventet. Sammenliknet med en person med folkeskole, vil en med framhaldsskole ligge ca. 17% over i inntekt, middel- realskole vil gi et tillegg på 36%, eksamen artium 37% og universitets- og høgskoleutdanning hele 84%. Det som kanskje kunne være verdt å merke seg er den relativt lille forskjellen i inntektstillegg mellom middelskole og eksamen artium. Man kan muligens si at det lønner seg

(inntektsmessig) mindre enn oftest antatt å ta eksamen artium uten å ta en universitetsutdanning. Konklusjonen kan svekkes en del når man tar i betraktning aldersfordelingen i materialet. (Jfr. bemerkninger om dette under I).

En signifikant forskjell i lønn i forhold til referansegruppen "primærnæringer og industri" viste gruppen "Kontor og handelsarbeide, teknisk, vitenskapelig og administrativt arbeide". Koeffisienten for gruppen "Service, transport, militært etc." ble ikke signifikant forskjellig fra 0, i overensstemmelse med resultatet fra I. Yrke var forventet å ha en god del betydning for lønnens høyde, men det kan være at vår modell ikke er særlig skikket til å vise slike forskjeller. I dette tilfellet er det antagelig fordi vi har aggregert yrkesgruppene så sterkt at forskjellene utviskes.

For b_8 og b_9 dvs. bebyggelse på hjemstedet og handikap/ikke handikap blir estimatene av samme størrelsesorden som i modell I, og også her signifikante.

For gruppen sosial status er resultatene signifikante og mer "fornuftige" enn i mod. I. Det eneste man kunne reflektere over var at gruppen "selvstendige yrkesaktive" har ca. 18% mindre inntekt enn ansatte yrkesaktive. Resultatet blir imidlertid mer rimelig når man tar i betraktning at en meget stor del av de selvstendige yrkesaktive er bønder og fiskere som jevnt over har lave inntekter. Man skal heller ikke se bort fra at det her kan spille en rolle at vårt inntektsmateriale skriver seg fra skatteligningens "antatt inntekt".

For den siste inndelingskategori "bosted under oppveksten", skiller resultatene seg lite ut fra de under modell I, og man kan anvende de tilsvarende betraktninger.

Konklusjonen på sammenlikningen av de to modeller må bli at modell 2 må være å foretrekke ved eventuelle videre undersøkelser. For det første viser det seg at de samspillvariable nesten alle like godt kan være lik 0. For det andre blir tolkningsprobleme større når man ser på modell I, og dette er jo et vesentlig moment hvis de estimerte koeffisienter skal kunne brukes til noe annet enn å forutsi grupper av individers inntekt på grunnlag av oppgaver over sosio-økonomiske karakteristika.

Aldersfaktoren

I tidligere analyser av samme art som her, har man sett på aldersfaktoren mer spesielt. Vi kan lett finne den alder som gir maksimum fortjeneste.

I modell II får vi:

$$\frac{\partial \ln Y}{\partial A} = a_2 + 2a_3 A \quad \text{og ved å sette dette lik 0}$$

$$A^{\max} = -\frac{a_2}{2a_3}$$

I modell I blir uttrykket litt mer komplisert når vi tar hensyn til sammensetningen av de samspillvariable.

$$\frac{\partial \ln Y}{\partial A} = a_2 + 2a_3 A + c_1 R_y + c_2 Q_6 + c_3 Q_7 + c_4 R_u + c_5 Q_2 + c_6 Q_3 + c_7 Q_4 + c_8 Q_5$$

og ved å sette lik 0 og løse:

$$A^{\max} = - \frac{a_2 + c_1 R_y + c_2 Q_6 + c_3 Q_7 + c_4 R_u + c_5 Q_2 + c_6 Q_3 + c_7 Q_4 + c_8 Q_5}{2a_3}$$

Annenordensbetingelsene for max ser på bakgrunn av estimatene ut til å være oppfylt for begge modellene.

Vi ser her klart virkningen av de samspillvariable, vi får trukket inn i analysen en antagelse om at alderen for toppinntekten forandrer seg etter hvilken utdannings- og yrkesklasse personen tilhører. Vi har i modellen antatt at det er de ovennevnte karakteristiska som ut fra vårt synspunkt er de mest interessante mht. inntektens forløp med alderen. Ved å danne variable som uttrykker alderens samspill med andre faktorer, f.eks. kjønn, kunne vi også studere hvordan aldersforløp varierte med disse karakteristika.

Nedenstående tabell viser den alder hvor man i følge våre estimer oppnår maksimumsinntekt for de forskjellige inntekts/utdanningsstyper.

	Primærnæring, industri	Service, trsp., militært, etc.	Kontor, vitensk., administrasjon
Folkeskole	39,2	35,3	45,1
Framhaldsskole	43,7	39,7	49,4
Middelskole	55,1	51,1	60,9
Eks. artium	69,1	65,1	73,3
Univ., Høgsk.	69,1	65,2	73,4

Modell II gir et inntektsmaksimum for 48,3

Det som straks slår en ved denne tabellen er de store forskjeller i inntekts-max-alder det er mellom utdanningsgruppene. Det er naturlig at det er de laveste utdanningsklassene som har den laveste maksimumsalder, disse kommer f.eks. tidligere ut i arbeidslivet. Men at det skulle være en differanse på ca. 30 år mellom den alder der en mann med folkeskole og en mann med universitetsutdanning når sitt inntektsmaksimum virker unektelig overraskende.

Det synes særlig å være mellom utdanningsnivåene middelskole og eksamen artium at forskjellen i alder for inntektsmaksimum er svært stor. At maksimum inntekt skulle være nådd på et alderstrian over den vanlige pensjonsalder synes urimelig. Men må imidlertid huske at materialet også omfatter selvstendige, og av disse kan det vel være endel som fortsetter sin virksomhet selv ut over

denne alder. Vi må videre ta i betraktning at vi også her har trukket inn ikke-signifikante beregningsresultater. Det er også en del forskjell mellom de forskjellige yrkesgrupper mht alderen for inntektsmaksimum, men ikke av størrelsesorden som for utdanningsgrupperingen. Man kan ikke se bort fra at det kan eksistere meget større forskjeller som er blitt skjult av den kraftige aggregeringen som er foretatt.

Interkorrelasjon

Vi nevnte under behandlingen av de statistiske forutsetninger for estimeringen problemet med interkorrelasjon i de forklaringsvariable. Etter beregning av korrelasjonsmatrisen kan det se ut som om dette problem ikke er så stort som først antatt. For å illustrere dette er nedenstående tabeller tatt med. De angir de funne enkle empiriske korrelasjonskoeffisientene mellom de egenskaper man a priori kunne vente var sterkest korrelert.

	Service, trsp. militært, etc.	Kontor, vitenskaplig, administrasjon
Framhaldsskole	0,01010	-0,01050
Middelskole	0,09112	0,17509
Eksamen artium	0,07610	0,31225
Univ. og høgsk.	-0,06303	0,32750

	Kvinne	Våningshus på gårdsbruk	Frittliggende enebolig
Framhaldsskole	-0,01244	0,04580	0,01748
Middelskole	-0,00653	-0,13997	0,07349
Eksamen artium	-0,00416	-0,12983	0,05828
Univ. og høgsk.	-0,05519	-0,10078	0,05233

LITTERATURLISTE

- (1) H.S. Houthakker: Education and Income. (Review of Economics and Statistics 1957)
- (2) F. Gerard Adams: The Size of Individual Incomes, Socio-Economic Variables and Chance Variation. (Review of Economics and Statistics 1958)
- (3) T.P. Hill: Analysis of the Distribution of Wages and Salaries in Great Britain. (Econometrica 1959)
- (4) J. Morgan og M. David: Education and Income. (Quarterly Journal of Economics 1963)
- (5) Martin Carnoy: Earnings and Schooling in Mexico. (Economic Development and Cultural Change, juli 1967)
- (6) P.A. Vatløe: Faktorer som påvirker inntektens størrelse for individer i Norge. (Arbeidsnotat fra SSB 21/8 1967)
- (7) Orientering for bruk av SSB's regresjonsprogram, 45 variable, dobbel presisjon. (Oslo 19/1 1968)
- (8) Herdis Thorén Amundsen: Innføring i Teoretisk Statistikk Bind III
- (9) Ivar Østby: Lønnsstrukturundersøkelser i varehandelen. (Arbeidsnotat fra SSB 23/12 1964)
- (10) Gardner Ackley: Macroeconomic Theory
- (11) J. Aitchison og J.A.C. Brown: The Lognormal Distribution
- (12) J. Johnston: Econometric Methods

V e d l e g g 1

BOLIGUNDERSØKELSEN OKTOBER 1967

Utdrag av skjema for husholdningsintervju

I. HUSHOLDNINGENS SAMMENSETNING M.V.

1. Først skulle jeg gjerne ha noen opplysninger om hver enkelt person i Deres husholdning. Til husholdningen regnes alle personer som bor fast i leiligheten og som har full kost eller i det minste middag, i husholdningen. Vi begynner med den såkalte hovedperson, det vil si den som skaffer til veie den største del av husholdningens inntekt.

Hush. med-lem nr, (6-7)	Navn	Fødselsdag og -år (8-13)	For Byrået (14-18)	Stilling i husholdningen (19-20)	Ekte-skape-lig status (21)	Hovedyrke. (For skoleelever oppgis skoleslag) S (selvst.), A (ansatt) (22)	Intervju (23) (24-25)
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							

3. Hva slags utdanning har hovedpersonen i husholdningen?

	Allmennutdanning	Spesialutdanning (før opp hva slag)
1	<input type="checkbox"/> Folkeskole	
2	<input type="checkbox"/> Framhaldsskole/ Ungdomsskole/ Folkehøgskole	
3	<input type="checkbox"/> Middelskole Realskole	
4	<input type="checkbox"/> Examen artium	

5.a. Er det noen i Deres husholdning som lider av varig legemlig sykdom eller handicap som i vesentlig grad hemmer arbeidsevne eller livsutfoldelse?

Nei \longrightarrow Gå til spn. 6.
 Ja

b. Hvem gjelder det og hva slag sykdom eller handicap dreier det seg om? Har vedkommende søkt eller fått noen ytelser fra det offentlige i form av uførtrygd, attføringspenger e.l.?

Hush.

medlem
nr.

Sykdom/handicap

Uførtrygd e.l.

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
		Ja		Nei		Vet ikke			
		"		"		" "			
		"		"		" "			

7. Hva slags hustype bodde De i under oppveksten?

- 2-14
- 1 Våningshus på gårdsbruk
 - 2 Frittliggende enebolig
 - 3 Kjedehus, rekkehus, inklusive vertikalt delte tomannsbolig
 - 4 To- til firemannsbolig, horisontalt delt
 - 5 Hus med 5 eller flere leiligheter (ikke rekke- eller kjedehus)
 - 6 Kombinert bolig- og forretningsbygg
 - 7 Annen hustype: _____
-

9.f. På hva slags sted ligger boligen?

- 2-21
- 1 I spredtbygde strøk
 - 2 I tettbygde strøk med 200-2000 innbyggere
 - 3 " " " " 2000-20000 "
 - 4 " " " " over 20000 "

V e d l e g g 2

BOLIGUNDERSØKELSEN OKTOBER 1967

Utdrag av kodeliste

HUSHOLDNINGSINTERVJUSpørsmål 1

STILLING I HUSHOLDNINGEN (Kol. 19-20)

- 01 = Hovedperson (Hp)
- 02 = Ektefelle (Ef)
- 03 = Barn
- 04 = Svigerinne, svoger
- 05 = Foreldre
- 06 = Svigerforeldre
- 07 = Søsken
- 08 = Andre familiemedlemmer, som barnebarn, svigerbarn, onkel, tante
- 09 = Losjerende med kost (L)
- 10 = Arbeidshjelp (Arb)
- 11 = Ikke kodbart

SOSIAL STATUS (Kol. 22)

Kodes til høyre i rubrikk 22.

- 1 = Selvstendig yrkesaktiv
- 2 = Ansatt yrkesaktiv
- 3 = Hjemnearbeidende husmødre
- 4 = Elev av folkeskole, barneskole
- 5 = Elev av framhaldsskole, realskole, ungdomsskole, folkehøgskole, yrkesskole
- 6 = Elev av gymnas og høyere skoler, student
- 7 = Hjemneværende barn
- 8 = Pensjonister, trygdede og andre ikke yrkesaktive
- 9 = Ikke kodbart

YRKE (Kol. 24)

Kodes bare for yrkesaktive (kode 1 og 2 under sosial status) etter Standard for yrkesgruppering. Sett strek ut for ikke-yrkesaktive personer av hensyn til punchingen.

- 0 = Teknisk, vitenskapelig, humanistisk og kunstnerisk arbeid
- 1 = Administrasjons- og forvaltningsarbeid, bedrifts- og organisasjonsledelse
- 2 = Kontorarbeid
- 3 = Handelsarbeid
- 4 = Jordbruks-, skogbruks- og fiskearbeid
- 5 = Gruve- og sprengningsarbeid m.m.
- 6 = Transport- og kommunikasjonsarbeid
- 7 = Industriarbeid, Bygge- og anleggsarbeid
- 8 = Servicearbeid og Militært arbeid
- 9 = Uoppgitt, ikke kodbart

TYPE SYKDOM/HANDIKAP (Kol. 25), ffr. spm. 5b

Kodes ut for hush.medl.nr. for den person som har handikap.

- 1 = Hjerte- og karsykdommer, inklusive hjerneslag
- 2 = Skjelett-, hud- og muskelsykdommer, misdannelser
- 3 = Kroniske lungesykdommer (lungetuberkulose, bronkitt, astma)
- 4 = Skader og amputasjoner som følge av ulykker
- 5 = Sykdommer i nervesystemet, sansedefekter, polio
- 6 = Allergiske, hormonelle, stoffskifte- og blodsykdommer, sykdommer i fordøyelses- og genitalorganer
- 7 = Sinnslidelser
- 8 = Andre sykdommer
- 9 = Uoppgitt

Uføretrygd E.L. (Kol. 26), kfr. spm. 5b

- 1 = Ja
- 0 = Nei
- 9 = Vet ikke

UTDANNING (Spm. 3 og 4, Personintervju spm. 1)

Koden skrives umiddelbart til høyre for puncheposisjonen.

- 0 = Bare 7-årig folkeskole
- 1 = 7-årig folkeskole med minst 6 måneders spesialutdanning
- 2 = Framhaldsskole m.v.
- 3 = Framhaldsskole m.v. med minst 6 måneders spesialutdanning
- 4 = Middelskole, realskole
- 5 = Middell-/realskole med minst 1 års spesialutdanning
- 6 = Eksamen artium
- 7 = Eksamen artium med minst 1 års spesialutdanning utenom universitetet eller høgskole
- 8 = Universitet eller høgskole
- 9 = Vet ikke, uoppgitt

ANTATT FORMUE (Kol. 27-29)

ANTATT INNTEKT (Kol. 30-32)

Antatt formue og inntekt til statsskattelikningen 1966 tas fra skatteprotokollene. Dersom formue/inntekt til statsskattelikningen ikke er ført opp, nyttes formue/inntekt ved kommuneskattelikningen. Dette arbeidet utføres som egen operasjon.

Antatt formue kodes i hele 1 000 kroner og med 3 siffer. Formue på mer enn 990 000 kroner, kodes 990.

Antatt inntekt kodes i hele 100 kroner og med 3 siffer. Inntekt høyere enn 99 000 kroner, kodes 990.

Kodingen foretas i høyre marg under spm. 1 ut for de personer som har formue/inntekt og på 2 linjer, formuen øverst, inntekten nedenfor.