

# Arbeidsnotater

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

IO 66/5

Oslo, 24. juni 1966

## Utvalgsplan for varehandel - regresjonsberegninger

Av

Eivind Gilje

### I n n h o l d

	Side
1. Innledning .....	1
2. Matematisk modell for regresjonsberegningene .....	3
3. Variable .....	6
4. Hva er gjort? .....	8
5. Tabeller:	
Tabell 1: Variansanalyser .....	9
Tabell 2: Forventninger, varianser og korrelasjons- koeffisinerter .....	10
Tabell 3: Regresjonskoeffisienter .....	13
Tabell 4: Regresjonskoeffisienter .....	16
Tabell 5: Kurveligninger .....	19
6. Tabellkommentarer og konklusjoner .....	21

## 1. Innledning

Dette arbeidet er direkte foranlediget av arbeidsnotat IO 64/4 av Egede Larssen, men som en vil se er de undersøkelsene som er gjort på enkelte punkter mer omfattende, på andre mindre enn det han ville ha klarlagt. Jeg vil derfor prøve å forme dette notatet slik at det kan leses og resultatene tolkes uten at man på forhånd kjenner Egede Larssens problemstilling.

Vi er interessert i å undersøke egenskapene ved en ny estimeringsmetode i varehandel og har først inndelt denne i tre hovedgrupper: Engros, detalj innen nærings- og nydelsesmidler og annen detalj. Hver hovedgruppe er igjen inndelt i undergrupper (strata). Hvilke strata vi har går tydelig frem av tabellene. Nå viser det seg at ved denne inndeling blir det totale antall strata stort og de enkelte strata ofte små. Spesielt gjør dette seg gjeldende når man som i registeret også inndeler bedriftene etter størrelse, definert ved totalomsetningen.

I slike små strata vil estimatene for undersøkte kjennetegn være beheftet med stor usikkerhet hvis vi bruker vanlige gjennomsnitt. Anta nå at det er høy korrelasjon mellom totalomsetning og det undersøkte kjennetegn. Hvis vi har et noenlunde bra estimat på totalomsetningen for en bedrift, skulle man ikke da ha en mulighet for på grunnlag av dette å finne et bra estimat også for det andre kjennetegnet? De problemene som reiser seg her har Egede Larssen behandlet utførlig i sitt notat, så jeg hopper rett på konklusjonene og vil først gi noen definisjoner.

Anta først at bedriftene er inndelt i omsetningsgrupper, slik at bedrifter med omtrent samme totalomsetning kommer i samme gruppe. En tenker seg nå en "bakenforliggende" uendelig populasjon og innfører følgende betegnelser:

	Totalt			I stratum nr. h		
	Antall enheter	Totalomsetning	Kjennetegn som skal undersøkes	Antall enheter	Totalomsetning	Kjennetegn som skal undersøkes
Uendelig populasjon	$\infty$	$\xi$	$\eta$	$\infty$	$\xi_h$	$\eta_h$
Endelig populasjon	$N$	$\bar{A}$	$\bar{B}$	$N_h$	$\bar{A}_h$	$\bar{B}_h$
Utvalg .....	$n$	$\bar{x}$	$\bar{y}$	$n_h$	$\bar{x}_h$	$\bar{y}_h$
Varianser .....		$\sigma^2$	$\tau^2$		$\sigma_h^2$	$\tau_h^2$

Vi er nå interessert i å finne et estimat for  $\bar{B}_h$ . Egede Larssen innfører de tre estimatorer  $\bar{y}_h$ ,  $\bar{y}$  og  $\bar{B}_h = \frac{\xi_h}{\xi} \bar{y}$  og sammenligner disse. Her er  $\bar{y}_h$  det vanlige gjennomsnittet innen den undersøkte undergruppe,  $\bar{y}$  kaller han det ukorrigerede gjennomsnitt og  $\bar{B}_h$  det korrigerede gjennomsnitt. Når det gjelder  $\bar{B}_h$ , må en være oppmerksom på at også  $\xi_h$  og  $\xi$  må estimeres. Jeg skal ikke komme nærmere inn på dette her.

Sammenligner en disse tre estimatorer under visse gitte forutsetninger som kommer tydelig frem av Egede Larssens notat, får en først for  $\bar{y}$  og  $\bar{B}_h$  :

Korrigerede gjennomsnitts metode gir et estimat med mindre varians enn ukorrigerede gjennomsnitts metode såfremt korrelasjonen mellom betingede forventninger gitt stratum for totalomsetningen og for kjennetegn som skal undersøkes, overstiger størrelsen  $\frac{C}{2}$ .

Her er:

$$C = \sqrt{\frac{E(\xi_h - \xi)^2}{E(\eta_h - \eta)^2}} \frac{\eta}{\xi} \quad (1.1)$$

Forventningene som inngår i (1.1) trenger en nærmere forklaring. Den stokastiske variable er her nemlig h. (Hvilket stratum). Dermed har vi

$$E(\xi_h - \xi)^2 = \sum_h (\xi_h - \xi)^2 \text{Pr}(h) \quad (1.2)$$

og tilsvarende for  $E(\eta_h - \eta)^2$ . Dette vil jeg komme tilbake til under tolkingen av resultatene.

Videre har vi:

For at ukorrigerede eller korrigerede gjennomsnitts metode skal være fordelaktig må under enhver omstendighet utvalget være på mindre enn 50 prosent. Er utvalget 50 prosent eller større, bør lineært estimat foretrekkes - dvs. gjennomsnitt pr. stratum.

Da jeg skulle foreta min analyse, viste det seg at mange av de beregningene jeg trengte kunne foretas på det eksisterende regresjonsprogram Byrået har. Selv om det setter en ganske vesentlig restriksjon på analysen, ble det besluttet å bruke dette programmet for å spare tid. Det viste seg nemlig at det var teknisk umulig både å dele opp materialet i omsetningsgrupper og strata. Dermed måtte oppdelingen i omsetningsgrupper falle.

Analysen som må ansees som foreløpig, er gjort sommeren 1965 på den del av materialet fra bedriftstellingene 1963, varehandel, som da var ferdig behandlet av maskinavdelingen. Dette materialet består av 11 384 bedrifter av en totalmasse på ca. 60 000. Her er altså foretatt et utvalg som ikke er tilfeldig trukket. Hvilken betydning dette har, vil jeg komme tilbake til senere, men her kan nevnes at det finnes strata som ikke er representert i det hele tatt i undersøkelsen.

Av denne og av grunner som vil komme klart frem senere, anbefales det å se på de fremkomne resultater i hvert fall når det gjelder regresjonskurvene kun som indikatorer på tendensen i materialet.

Bearbeidelsen er gjort på den elektroniske regnemaskin på Kjeller, og programmet er beskrevet i arbeidsnotat IB 64|5: "Multiple Regression and Correlation Analysis" av Svein Nordbotten og Thor Aastorp. Angående regresjonsanalysen henvises forøvrig til Statistisk Sentralbyrås håndbok nr. 22: "Statistisk testing av hypoteser ved regresjonsberegninger" av Harold W. Watts.

## 2. Matematisk modell for regresjonsberegningene

Som det fremgår av arbeidsnotatet til Egede Larssen kan en anta at totalomsetningen er logaritmisk- normalt fordelt. La oss i det følgende kalle totalomsetningen for  $X$  og det undersøkte kjennetegn for  $Y$ . Hva kan vi nå si om fordelingen til  $Y$ ? I grunnen ingenting ut ifra det vi vet a priori, men hvis vi forlanger en forholdsvis høy korrelasjon mellom  $X$  og  $Y$ , kan vi i hvert fall foreløpig anta at også  $Y$  er logaritmisk-normal. Alt som sies her gjelder foreløpig kun en bestemt under- alternativt hovedgruppe. Vi vil videre anta at observasjonene  $Y_1, Y_2, \dots, Y_n$  hvor  $n$  er antall bedrifter, er stokastisk uavhengige. Nå er vi interessert i å finne den eventuelle sammenheng mellom  $E(Y|X=x)$  og  $x$ . Denne kan settes:

$$E(Y|X=x) = f(x) \quad (2.1)$$

Å finne denne funksjonen er så godt som umulig og kan like godt oppgis, men vi kan gjøre noe annet. Vi kan velge en funksjon av  $x$  og prøve å tilpasse denne til  $E(Y|X=x)$ . Vi velger altså funksjonsformen og tilpasser de ukjente parametrene.

Å finne en passende funksjonsform er ikke lett. For det første bør den være så robust at enhver rimelig sammenheng mellom  $Y$  og  $X$  kan beskrives, men den bør også være så enkel at det ikke byr på for store

vansker å tolke det fremkomne resultat. For det andre vil vi gjerne ha en lineærsammenheng med hensyn på de ukjente parametrene mellom  $\ln Y$  og  $\ln f(x)$ . Grunnen til dette skal forklares om et øyeblikk. Vi må videre la muligheten for å teste en eventuell lineær sammenheng mellom  $Y$  og  $X$  stå åpen. Dette p.g.a. den store interesse et slikt resultat ville ha.

Da  $X$  og  $Y$  er antatt logaritmisk-normalt fordelte, vil altså  $\ln X$  og  $\ln Y$  være normalfordelte. Dette leder oss under de forutsetningene vi har gjort til en vanlig lineær regresjonsmodell med normalfordelte variable. Vi antar altså sammenhengen:

$$\ln Y_i = \ln f(x_i) + W \quad (2.2)$$

der  $W$  er en variabel som er  $N(0, \tau)$  hvor  $\text{var } \ln Y_i = \tau^2$ .

En funksjonsform som tilfredsstillere alle de nevnte forutsetninger unntatt kanskje den om robusthet, er:

$$f(x_i) = \mu x_i^\gamma e^{\beta x_i^{-2}} \quad (2.3)$$

eller på logaritmisk form:

$$\ln Y_i = \ln \mu + \gamma \ln x_i + \beta \frac{1}{x_i^2} + W \quad (2.4)$$

hvor  $\mu$ ,  $\gamma$  og  $\beta$  er ukjente parametre som skal estimeres. Som en vil se er det her komme med en ekstra forutsetning;  $E(Y|X=x)$  vil alltid gå gjennom origo. Dette har sannsynligvis liten eller ingen betydning for det fremkomne resultat. Det betyr jo bare at  $Y$  (f.eks. bruttoinntekten) er lik null for omsetningen lik null; noe som vel må antas rimelig.

Hvordan vil nå (2.3) se ut grafisk? Vi deriverer to ganger.

$$\frac{dY_i}{dx_i} = \mu x_i^{\gamma-1} e^{\beta x_i^{-2}} (\gamma - 2\beta x_i^{-2}) \quad (2.5)$$

$$\frac{d^2Y_i}{dx_i^2} = \mu x_i^{\gamma-2} e^{\beta x_i^{-2}} \left[ (\gamma - 2\beta x_i^{-2})^2 - \gamma + 6\beta x_i^{-2} \right] \quad (2.6)$$

Vi vet a priori at kurven er ikke-synkende. Ergo må  $\mu \geq 0$  og  $(\gamma - 2\beta x_i^{-2}) \geq 0$ . Begge ledd kan for såvidt også være negative, men da må  $\beta$  være så stor at den alltid er større eller ubetydelig mindre enn  $x_i^2$  noe som er høyst urealistisk. Av (2.6) ser vi at kurven er konveks hvis  $(\gamma - 2\beta x_i^{-2})^2 \geq \gamma - 6\beta x_i^{-2}$  og konkav hvis  $(\gamma - 2\beta x_i^{-2})^2 \leq \gamma - 6\beta x_i^{-2}$ .

Den har et vendepunkt hvis likhetstegn gjelder. For forholdsvis store  $x_i$  vil dette si at kurven er konveks hvis  $\gamma$  er større enn 1 og konkav hvis det motsatte gjelder. For likhet får vi en rett linje.

Kort kan vi nå si at kurven tilnærmet får form av en sammensetning av to parabler med omventet krumning. Dette fordi den kun kan ha ett vendepunkt. Dette setter nokså sterke restriksjoner på modellen, men tenker vi etter på hva vi intuitivt vet om de størrelser som inngår, er restriksjonene ikke så urimelige.

- - -

Vi har til nå bare behandlet funksjonsformen innen en bestemt underalternativt hovedgruppe. Hva om vi nå forsøker å omforme modellen slik at vi greier oss med en regresjonsanalyse i stedet for en for hver gruppe? Dette kan gjøres ved å innføre såkalte "dummy"-variable. Dvs. variable som kun kan anta verdiene 0 eller 1. Vi får da i (2.4) k ekstra ledd hvis vi har  $k + 1$  undergrupper:

$$\ln Y_i = \ln \mu + \gamma \ln x_i + \beta x_i^{-2} + \sum_{j=1}^k \alpha_j Q_j + W \quad (2.7)$$

Vi har altså her en referansegruppe for hvilken alle  $Q_j = 0$ , og  $Q_j$  antar verdien 1 hvis en observasjon kommer fra den  $i$ -te av de  $k$  andre undergrupper; 0 ellers.

Det er klart at dette siste setter enda større restriksjoner på modellen. Vi får riktignok en regresjonskurve for hver undergruppe, men alle kurvene er "avhengige" av hverandre. Matematisk vil dette si at vi faktisk mister en dimensjon av variasjonsområdet for regresjonskurvene; nemlig variasjonen i krumningen. Krumningen blir fastsatt en gang for alle og variasjonen fra kurve til kurve blir bare bestemt av tangentens helningsvinkel i den samme avstand fra origo.

Den kurven vi får frem for undergruppe  $j$  blir:

$$E(Y|X = x, Q_j = 1, Q_i \neq j = 0) \quad \text{for alle } i \quad (2.8)$$

og ikke

$$E(Y|X = x, Q_j = 1) \quad (2.9)$$

som er den som egentlig har interesse selv om (2.8) er tilnærmet lik denne. Det kan synes som om de to kurver skal bli helt like fordi hvis  $Q_j$  er gitt lik 1, må alle  $Q_{i \neq j}$  være lik 0, men dette er ikke tilfelle da i (2.8) alle undergrupper er med på å bestemme forventningsverdien, mens i (2.7) kun undergruppe  $j$  er med.

Her kommer vi inn på en av fordelene med den utvidede modell. Vi vil i undergrupper med få observasjoner få bedre estimater på regresjonskurvene her enn i den første modellen. Dette nettopp fordi observasjonene fra alle undergrupper er med på å estimere kurven for hver enkelt undergruppe. Her er selvsagt forutsatt at regresjonskurvene ikke er totalt forskjellige fra hverandre.

Den andre og vel den største fordelen er at vi nå greier oss med én regresjonsanalyse, vel å merke på langt flere variable.

Av disse to grunner er den utvidede modell valgt i denne foreløpige undersøkelse.

### 3. Variable

De variable som er brukt i analysen er enten de direkte opplysningene som kommer fra bedriftstellingen 1963, eller transformasjoner av disse. Transformasjonene er av tekniske grunner foretatt i to omganger. I første omgang er følgende transformasjoner foretatt:

- I. Eiere og ansatte er summert og har fått navnet sysselsatte.
- II. Omsetning for egen regning i engros og detalj er summert og har fått navnet omsetning.
- III. Bruttoinntekten er en lineærtransformasjon av følgende variable: omsetningen (S), omsetningsavgiften (A) som igjen er summen av 10 prosent-avgiften og andre avgifter, varekjøp (I), varebeholdning pr. 31/12-63 ( $L_2$ ) og varebeholdning pr. 1/1-63 ( $L_1$ ). Transformasjonen har følgende form hvis bruttoinntekten kalles for B:

$$B = (S - A) - (I - (L_2 - L_1)) \quad (3.1)$$

Under selve regresjonsanalysen vil de variable så automatisk bli transformert over i den form de skal ha under regnearbeidet. På neste side følger en liste over de variable i denne form. De variable med nummer fra og med 12 til og med 49 er "dummyer" og gir, hvis de har verdien 1, den varehandelsgruppe som står til høyre på listen. Tallene som står foran varehandelsgruppen er fra "Standard for næringsgruppering i offentlig norsk statistikk", Statistisk Sentralbyrås håndbok nr. 9. Vår referansegruppe (alle  $Q_i = 0$ ) er kommisjonshandel, gruppering nr. 660.

<u>Variabel nr.:</u>	<u>Variabel:</u>	
1	ln Bruttoinntekten	
2	ln Sysselsatte	
3	ln Omsetningen	
4	1 : (omsetningen) <sup>2</sup>	
5	Varebeh. pr. 1/1-63	
6	ln (Varebeh. pr. 31/12-63)	
7	Personbiler	
8	Busser	
9	Lastebiler	
10	ln Bygningsforsikring	
11	ln Inventarforsikring	
Engros	12	621. Nærings- og nytelsesmidler 1
	13	622. Nærings- og nytelsesmidler 2
	14	623. Tekstil og bekledning
	15	625. Ur, optiske artikler, musikkinstrumenter, m.v.
	16	626. Sykepleie og apotekervarer
	17	627. Papir, leketøy og kortevarer
	18	628. Gull- og sølvvarer, porteføljevarer, kosmetikk
	19	631. Møbler og innbo
	20	632. Jern- og fargevarer m.v.
	21	633. Bygningsartikler m.v.
	22	634. Biler og rekvisita
	23	635. Maskiner og utstyr
	24	636. Diverse råvarer
	25	637. Brensel og smøreoljer
	26	6391. Skrap, filler m.v.
	27	6392 og 6399. Levende dyr og Andre varer ikke nevnt før
	Detalj, nærings- og nydelses- midler	28
29		6412. Landhandel
30		6413, 6427 og 6429. Kaffe og te, Øl og mineralvann, Andre nærings- og nytelsesmidler
31		6414. Bakervarer
32		6415. Melk og delikatesser
33		6416. Kjøtt og pølsevarer
34		6417. Fisk og vilt
35		6421. Blomster
36		6418 og 6422. Frukt og grønnsaker og Frukt, grønnsaker, blomster
37		6423, 6424, 6425 og 6426. Tobakkvarer, Tobakk, sjokolade og frukt, Sjokolade og Sjokolade og frukt
38		6428. Vin og spiritiosa



<u>Variabel nr.:</u>	<u>Variabel:</u>
{ 39	643. Tekstil og bekledning 1
{ 40	644. Tekstil og bekledning 2
{ 41	645. Ur, optiske artikler, musikkinstrumenter m.v.
{ 42	646. Sykepleie og apotekerverer
{ 43	647. Bøker, papir og korteverer
{ 44	648. Gull og sølvvarer
{ 45	651. Møbler og innbo
{ 46	652. Jern- og fargevarer
{ 47	654. Biler og rekvisita
{ 48	657. Brensel og smøreoljer
{ 49	659. Detalj ellers

#### 4. Hva er gjort?

Maskinen skriver før den begynner på selve regresjonsanalysen, ut følgende:

- I. Summer av de variable
- II. Kvadratsummer av de variable
- III. Summer av kryssproduktene av de variable
- IV. Summer av kryssproduktene av avvikene fra middelverdiene
- V. Estimerer for korrelasjonskoeffisientene.

(De variable er her hele tiden de transformerte variable).

På disse er det så tilsammen gjort 13 forskjellige regresjonsanalyser. I 4 analyser er ln til bruttoinntekten tatt som avhengig variabel, i 4 andre ln til antall sysselsatte, i 2 varebeholdning pr. 1/1-63, i 2 ln til varebeholdning pr. 31/12-63 og i én er omsetningen den avhengige variabel. Alle "dummyvariable" er med i hver analyse. Nedenfor følger en liste over de analyser som er gjort. Til forklaringsvariablene på høyre side legges altså alle "dummyvariable".

<u>Analyse nr.:</u>	<u>Avhengig variabel:</u>	<u>Forklaringsvariable:</u>					
1	1	2	3	4			
2	1	3	4				
3	1	2	3				
4	1	3					
5	2	1	3	4			
6	2	3	4				
7	2	1	3				
8	2	3					
9	3	1	2				
10	5	2	3	4	10	11	
11	5	2	3	4			
12	6	2	3	4	10	11	
13	6	2	3	4			

Utenom den vanlige regresjons- og korrelasjonsanalysen foretas også en variansanalyse på det undersøkte materialets tilpasning til regresjonskurven.

Videre er det foretatt en ekstra kjøring for å finne varianser og covarianser mellom de tre variable bruttoinntekt, antall sysselsatte og omsetning (ikke logaritmer her) innen hver undergruppe. Dette kan gjøres meget enkelt ved å innføre alle mulige kryssprodukter av de tre variable som nye variable. Når så maskinen regner ut summer av produktene av disse kryssprodukter med "dummyvariablene", får vi ut summene av kryssproduktene innen hver undergruppe; og det er jo det som trengs for å regne ut de covariansmatriser vi er interessert i.

## 5. Tabeller

Tabell 1: Variansanalyser.  $f_1$  er antall frihetsgrader i telleren, og  $f_2$  er antall frihetsgrader i nevneren

Analyse nr.	$f_1$	$f_2$	F-verdi
1	39	11344	1987,82164
2	38	11345	1889,49195
3	38	11345	2032,44638
4	37	11346	1917,74692
5	39	11344	535,54830
6	38	11345	494,99159
7	38	11345	530,01178
8	37	11346	484,02307
9	38	11345	2525,48551
10	41	11342	57,30403
11	39	11344	59,84285
12	41	11342	621,28320
13	39	11344	628,45639

Tabell 2: Estimer for forventninger, varianser og korrelasjonskoeffisienter for de opprinnelige variable

Gruppe	Antall	Bruttoinntekt = X · 1000 kr.		Sysselsatte = Y		Omsetning = Z · 1000 kr.		$\rho(X,Y)$	$\rho(X,Z)$	$\rho(Y,Z)$
		$\bar{X}$	var X	$\bar{Y}$	var Y	$\bar{Z}$	var Z			
Nærings- og nydelsesmidl. 1	87	346,1494	209140,72	7,39	54,01	2612,299	$9,2664 \cdot 10^6$	0,684	0,821	0,73
M. og n. 2	50	186,4800	74696,87	4,86	18,12	1428,440	$2,7768 \cdot 10^6$	0,549	0,877	0,59
Tekstil og bekl.	104	311,2596	146407,48	6,94	47,22	1750,904	$4,1026 \cdot 10^6$	0,755	0,908	0,85
Ur, opt., musikk	26	366,3462	275032,70	6,58	57,93	1738,192	$6,8416 \cdot 10^6$	0,944	0,971	0,89
Sykepl. og apot.	0	-	-	-	-	-	-	-	-	-
Pap., leketøy og kortev.	41	170,8049	79239,31	5,68	67,02	801,561	$1,6954 \cdot 10^6$	0,986	0,982	0,99
Gull og sølv, portefølje og kosmetikk	29	410,3793	499180,17	7,76	86,26	1757,069	$8,3067 \cdot 10^6$	0,911	0,967	0,92
Möbler og innbo	45	209,4667	72365,25	4,00	13,59	1659,600	$7,9197 \cdot 10^6$	0,696	0,841	0,49
Jern- og fargev.	43	302,0930	125269,94	8,33	89,65	1823,395	$5,5549 \cdot 10^6$	0,947	0,922	0,78
Bygn. art.	148	268,6757	218554,45	6,77	92,23	1976,669	$18,2703 \cdot 10^6$	0,899	0,855	0,88
Biler og rekv.	21	1183,5238	12844420,5	12,90	668,69	11634,904	$1884,5812 \cdot 10^6$	0,985	0,998	0,98
Mask. og utstyr	285	245,3895	164225,29	5,89	66,77	1352,751	$6,0234 \cdot 10^6$	0,902	0,947	0,33
Div. råvarer	46	219,8043	236751,32	4,17	40,50	1282,935	$8,0269 \cdot 10^6$	0,552	0,615	0,82
Brensel og smøreoljer	12	449,6667	118947,15	9,67	54,61	2153,167	$3,1529 \cdot 10^6$	0,928	0,963	0,85
Skrap, filler	98	70,5612	18554,70	2,56	9,69	208,020	$0,1420 \cdot 10^6$	0,881	0,915	0,74
Lev. dyr og andre varer	39	97,3590	15751,39	3,03	13,18	575,795	$1,5460 \cdot 10^6$	0,783	0,530	0,23
Engros	1074	263,8994	414620,62	5,96	69,72	1688,077	$44,1488 \cdot 10^6$	0,804	0,931	0,74

(forts.)

Tabell 2 (forts.)

Gruppe	Antall	Bruttoinntekt = $\bar{X} \cdot 1000$ kr.		Sysselsatte = Y		Omsetning = Z · 1000 kr.		$\rho(X, Y)$	$\rho(X, Z)$	$\rho(Y, Z)$
		$\bar{X}$	var X	$\bar{Y}$	var Y	$\bar{Z}$	var Z			
Kolonialv.	3697	53,2256	3068,34	3,16	5,51	400,313	132096,40	0,866	0,952	0,88
Landhandel	3665	44,9250	1882,86	2,57	2,72	369,723	95795,53	0,802	0,898	0,85
Kaffe og te, Öl og mineralvann, Andre n. og n.	20	40,4500	1630,26	2,20	2,06	236,100	62209,78	0,849	0,974	0,78
Bakerv.	65	26,7538	388,97	1,91	1,71	139,169	10462,61	0,704	0,862	0,81
Melk og delikat.	125	33,2880	626,30	2,12	1,56	242,072	27047,18	0,806	0,912	0,72
Kjøtt og pølsev.	213	113,4930	11817,48	5,81	26,51	547,596	196864,98	0,806	0,944	0,85
Fisk og vilt	122	46,0820	2525,35	2,34	3,24	233,959	50588,62	0,755	0,887	0,80
Blomster	183	45,7760	4792,58	2,77	9,45	144,672	36109,66	0,960	0,970	0,85
Frukt og grønnsaker	44	49,2045	2128,26	2,80	3,56	215,068	35057,04	0,745	0,793	0,68
Tobakk og sjokolade	634	28,1775	653,63	2,00	1,76	197,893	24652,93	0,782	0,928	0,68
Vin og spiri- tuosa	0	-	-	-	-	-	-	-	-	-
Detalj, n. og n.	8968	48,2869	2705,70	2,83	4,83	359,632	108534,49	0,847	0,906	0,82
Tekstil og bek. 1	18	48,5556	2479,91	2,17	4,85	244,722	71098,57	0,900	0,951	0,98
T. og b. 2	213	89,1737	10917,25	3,40	6,21	475,052	148737,59	0,698	0,760	0,90
Ur, opt., musikk	289	69,3668	6969,65	2,63	4,62	207,799	67909,58	0,866	0,944	0,88
Sykepl. og apot.	1	204,0000	-	7,00	-	529,000	-	-	-	-
Bøker, pap., kort.	6	116,5000	14121,90	5,33	13,07	494,500	142644,70	0,857	0,989	0,90
Gull og sølv	6	49,6667	875,47	2,00	1,20	241,667	19899,07	0,953	0,978	0,82

(forts.)

Tabell 2 (forts.)

Gruppe	Antall	Bruttoinntekt = X · 1000 kr.		Sysselsatte = Y		Omsetning = Z · 1000 kr.		$\rho(X, Y)$	$\rho(X, Z)$	$\rho(Y, Z)$
		$\bar{X}$	var X	$\bar{Y}$	var Y	$\bar{Z}$	var Z			
Möbler og innb.	454	97,4912	14023,43	3,36	10,16	503,863	292882,78	0,862	0,879	0,76
Jern- og fargev.	10	45,9000	2132,77	2,30	2,23	241,500	45267,39	0,750	0,887	0,87
Biler og rekv.	29	58,2759	4172,06	1,83	1,29	340,897	92025,81	0,689	0,780	0,57
Brensel og smøreoljer	44	26,0227	576,44	1,34	0,97	274,523	59782,91	0,519	0,663	0,70
Detalj ellers	25	46,8400	16757,31	1,92	6,74	202,880	245331,94	0,981	0,997	0,94
Detalj utenom n. og n.	1095	82,0475	10755,98	3,00	7,37	391,595	197986,93	0,834	0,849	0,82
Total	11384	70,8950	46281,18	3,12	11,99	480,975	4,4210 · 10 <sup>6</sup>	0,758	0,920	0,63
Total	11384	$\overline{\ln X}$	var ln X	$\overline{\ln Y}$	var ln Y	$\overline{\ln Z}$	var ln Z	$\rho(\ln X, \ln Y)$	$\rho(\ln X, \ln Z)$	$\rho(\ln Y, \ln Z)$
		3,5251	1,5121	0,87997	0,4540	5,4860	1,6911	0,740	0,895	0,679

I totalene er kommisjonshandel tatt med. Det er 247 bedrifter med i denne gruppe.

Tabell 3. Regresjonskoeffisienter etc. med ln til bruttoinntekten som avhengig variabel

Gruppe	Analyse nr. 1			Analyse nr. 2			Analyse nr. 3			Analyse nr. 4		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Nærings- og nydelsesmidl. 1	-0,77093	0,06474	-11,9082	-1,35831	0,06331	-21,4550	-0,71166	0,06413	-11,0968	-1,29075	0,06332	-20,3834
N. og n. 2	-0,79261	0,07579	-10,4583	-1,40476	0,07501	-18,7268	-0,73589	0,07536	-9,7657	-1,34361	0,07518	-17,8714
Tekstil og bekl.	-0,46302	0,06073	-7,6238	-1,01408	0,05939	-17,0743	-0,40744	0,06016	-6,7721	-0,95072	0,05940	-16,0040
Ur, opt., musikk	-0,31632	0,09593	-3,2973	-0,87105	0,09704	-8,9766	-0,26065	0,09567	-2,7245	-0,80779	0,09735	-8,2975
Sykepl. og apot.	-	-	-	-	-	-	-	-	-	-	-	-
Pap., leketøy og kortev.	-0,43503	0,07939	-5,4795	-0,93667	0,07996	-11,7146	-0,38579	0,07912	-4,8758	-0,88149	0,08020	-10,9907
Gull og sølv, portefølje og kosmetikk	-0,31319	0,09137	-3,4278	-0,81204	0,09264	-8,7654	-0,26201	0,09114	-2,8748	-0,75311	0,09295	-8,1019
Möbler og innbo	-0,49350	0,07831	-6,3015	-1,10307	0,07777	-14,1846	-0,41534	0,07742	-5,3650	-1,00235	0,07760	-12,9167
Jern- og fargev.	-0,50004	0,07863	-6,3594	-1,01140	0,07906	-12,7925	-0,44775	0,07830	-5,7182	-0,95130	0,07927	-12,0007
Bygn.art.	-0,68565	0,05534	-12,3907	-1,23560	0,05347	-23,1065	-0,62912	0,05467	-11,5084	-1,17042	0,05341	-21,9157
Biler og rekv.	-0,29393	0,10522	-2,7934	-0,82335	0,10701	-7,6942	-0,23775	0,10500	-2,2643	-0,75738	0,10738	-7,0532
Mask. og utstyr	-0,44294	0,04905	-9,0298	-0,99326	0,04645	-21,3819	-0,38555	0,04825	-7,9911	-0,92653	0,04628	-20,0211
Div. råvarer	-0,49042	0,07659	-6,4033	-1,06040	0,07632	-13,8933	-0,43956	0,07627	-5,7631	-1,00705	0,07655	-13,1547
Brensel og smørreoljer	-0,35133	0,13353	-2,6310	-0,80814	0,13702	-5,8981	-0,29955	0,13349	-2,2440	-0,74514	0,13759	-5,4155
Skrap, filler	-0,03181	0,05733	-0,5549	-0,49886	0,05670	-8,7983	0,03305	0,05646	0,5853	-0,41257	0,05642	-7,3119
Lev. dyr og andre varer	-0,35775	0,08045	-4,4469	-0,89778	0,08074	-11,1192	-0,30781	0,09114	-2,8748	-0,84401	0,08100	-10,4199
Engros	-0,47620	-	-	-1,02162	-	-	-0,41879	-	-	-0,95451	-	-

Tabell 3 (forts.)

Gruppe	Analyse nr. 1			Analyse nr. 2			Analyse nr. 3			Analyse nr. 4		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Kolonialv.	-0,39926	0,04072	-22,0857	-1,45098	0,03682	-39,4043	-0,85011	0,04000	-21,2527	-1,39948	0,03670	-38,1321
Landhandel	-0,97040	0,04119	-23,5598	-1,55934	0,03660	-42,6105	-0,92043	0,04046	-22,7516	-1,50896	0,03648	-41,3601
Kaffe og te, Öl og mineralvann, Andre n. og n.	-0,57099	0,10586	-5,4696	-1,14366	0,10743	-10,6456	-0,53165	0,10575	-5,0272	-1,09640	0,10789	-10,1622
Bakerov.	-0,53557	0,06647	-8,0573	-1,08237	0,06568	-16,4799	-0,49262	0,06621	-7,4398	-1,04193	0,06591	-15,8092
Mølk og dolikat.	-0,81863	0,05598	-14,6224	-1,40657	0,05364	-26,2210	-0,77090	0,05554	-13,8797	-1,36023	0,05374	-25,3090
Kjøtt og pølsev.	-0,53071	0,04943	-10,7364	-0,97897	0,04834	-20,2509	-0,48379	0,04893	-9,8883	-0,92431	0,04832	-19,1276
Fisk og vilt	-0,53974	0,05565	-9,6984	-1,09694	0,05372	-20,4191	-0,49420	0,05525	-8,9441	-1,05247	0,05384	-19,5489
Blomster	-0,21316	0,04845	-4,3997	-0,66253	0,04725	-14,0208	-0,17472	0,04813	-3,6303	-0,62351	0,04736	-13,1663
Frukt og grønnsaker	-0,46801	0,07606	-6,1530	-0,94711	0,07662	-12,3619	-0,42591	0,07588	-5,6129	-0,90347	0,07690	-11,7486
Tobakk og sjokolade	-0,82822	0,04160	-19,9116	-1,38761	0,03770	-36,8035	-0,78018	0,04093	-19,0609	-1,33869	0,03762	-35,5821
Vin og spirituosa	-	-	-	-	-	-	-	-	-	-	-	-
Detalj, n. og n.	-0,88749	-	-	-1,45081	-	-	-0,83853	-	-	-1,40048	-	-
Tekstil og bek. 1	-0,55202	0,11111	-4,9684	-1,15344	0,11270	-10,2350	-0,50456	0,11102	-4,5446	-1,10854	0,11319	-9,7933
T. og b. 2	-0,57693	0,05038	-11,4526	-1,13107	0,04788	-23,6219	-0,52711	0,04981	-10,5824	-1,07851	0,04788	-22,5257
Ur, opt., musikk	-0,00564	0,04553	-0,1238	-0,50589	0,04329	-11,6865	0,04085	0,04498	0,9081	-0,45568	0,04326	-10,5337
Sykepl. og apot.	0,07560	0,44178	0,1711	-0,28188	0,45654	-0,6174	0,12044	0,44245	0,2722	-0,22466	0,45884	-0,4896
Bøker, pap., kort.	-0,79204	0,18380	-4,3093	-1,22753	0,18934	-6,4833	-0,74621	0,18395	-4,0566	-1,17397	0,19024	-6,1711
Gull og sølv	-0,39648	0,18409	-2,1537	-0,99827	0,18901	-5,2815	-0,34814	0,18422	-1,8898	-0,95180	0,18993	-5,0113

Tabell 3 (forts.)

Gruppe	Analyse nr. 1			Analyse nr. 2			Analyse nr. 3			Analyse nr. 4		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Möbler og innb.	-0,46529	0,04523	-10,2882	-1,02985	0,04184	-24,6148	-0,41167	0,04446	-9,2592	-0,97106	0,04170	-23,2886
Jern- og fargev.	-0,54088	0,14430	-3,7482	-1,07118	0,14789	-7,2430	-0,49678	0,14436	-3,4412	-1,02746	0,14859	-6,9146
Biler og rekv.	-0,40107	0,09172	-5,2451	-1,14253	0,09161	-12,4712	-0,42918	0,09148	-4,6914	-1,09373	0,09197	-11,8923
Bronsel og smøroljer	-1,21302	0,07875	-15,4132	-1,92609	0,07704	-25,0009	-1,16050	0,07841	-14,8009	-1,87824	0,07731	-24,2958
Detailj ellers	-0,50444	0,09548	-5,2834	-1,03448	0,09674	-10,6933	-0,45914	0,09535	-4,8153	-0,98854	0,09714	-10,1763
Detailj utenom n. og n.	-0,40012	-	-	-0,95299	-	-	-0,39048	-	-	-0,89914	-	-
<u>Variabel</u>												
ln Omsetningen	0,81375	0,00647	125,7835	0,95091	0,00436	218,0272	0,80257	0,00622	128,9916	0,94000	0,00427	220,3904
l : Omsetn. <sup>2</sup>	0,89042	0,14521	6,1872	1,60084	0,14786	10,8269						
ln Sysselsatte	0,27906	0,00998	27,9631				0,28976	0,00985	29,4296			
Konstantledd		-0,40359			-0,36314			-0,39906			-0,35196	

I kolonne merket "Std. av." står estimatet for standardavviket på regresjonskoeffisienten.

Kolonnen merket "t-verdi" brukes til å teste hypotesen om regresjonskoeffisienten er lik 0 eller ikke. Studentfordelingen har n - s frihetsgrader der n er totalt antall observasjoner, og s er antall variable som inngår i regresjonen.



Tabell 4: Regresjonskoeffisienter etc. med ln til antall sysselsatte som avhengig variabel

Gruppe	Analyse nr. 5			Analyse nr. 6			Analyse nr. 7			Analyse nr. 8		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	T-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Nærings- og nydelsesmidl. 1	-1,79004	0,05684	-31,5070	-2,10469	0,05761	-36,5359	-1,68252	0,05712	-29,4564	-1,99847	0,05820	-34,3376
N. og n. 2	-1,86885	0,06703	-27,8794	-2,19343	0,06826	-32,1358	-1,76842	0,06754	-26,1838	-2,09730	0,06910	-30,3516
Tekstil og bekl.	-1,74022	0,05294	-32,8719	-1,97453	0,05404	-36,5375	-1,64221	0,05322	-30,8568	-1,87492	0,05460	-34,3398
Ur, opt., musikk	-1,78645	0,08571	-20,8440	-1,98771	0,08829	-22,5126	-1,69053	0,08651	-19,5410	-1,88825	0,08948	-21,1029
Sykopl. og apot.	-	-	-	-	-	-	-	-	-	-	-	-
Pap., leketøy og kortev.	-1,58103	0,07080	-22,3320	-1,79746	0,07275	-24,7059	-1,49493	0,07143	-20,9278	-1,71069	0,07371	-23,2071
Gull og sølv, portefølje og kosmetikk	-1,59983	0,08181	-19,5550	-1,78746	0,08430	-21,2047	-1,51046	0,08259	-18,2886	-1,69480	0,08543	-19,8375
Möbler og innbo	-1,92933	0,06905	-27,9421	-2,18420	0,07076	-30,8680	-1,78050	0,06925	-25,7095	-2,02585	0,07132	-28,4036
Jern- og fargev.	-1,59860	0,07008	-22,8097	-1,83230	0,07194	-25,4700	-1,50495	0,07067	-21,2944	-1,73780	0,07286	-23,8522
Bygn.art.	-1,68506	0,04816	-34,9900	-1,97055	0,04866	-40,4994	-1,58158	0,04831	-32,7414	-1,86807	0,04908	-38,0580
Biler og rekv.	-1,70676	0,09443	-18,0751	-1,89701	0,09737	-19,4828	-1,60790	0,09534	-16,8645	-1,79329	0,09869	-18,1702
Mask. og utstyr	-1,74238	0,04170	-41,7835	-1,97188	0,04227	-46,6516	-1,64018	0,04172	-39,3162	-1,86698	0,04253	-43,8936
Div. råvarer	-1,79733	0,06774	-26,5314	-2,04235	0,06945	-29,4083	-1,71197	0,06834	-25,0515	-1,95847	0,07036	-27,8346
Brensel og smøreljer	-1,45010	0,12078	-12,0065	-1,63683	0,12467	-13,1289	-1,35538	0,12206	-11,1045	-1,53778	0,12646	-12,1600
Skrap, fillor	-1,55825	0,05007	-31,1198	-1,67352	0,05159	-32,4378	-1,43686	0,05011	-28,6764	-1,53784	0,05186	-29,6543
Lev. dyr og andre varer	-1,72758	0,07145	-24,1795	-1,93503	0,07347	-26,3386	-1,64389	0,07210	-22,7988	-1,85049	0,07445	-24,8563
Engros	-1,71827	-	-	-1,95433	-	-	-1,61518	-	-	-1,84882	-	-

Tabell 4 (forts.)

Gruppe	Analyse nr. 5			Analyse nr. 6			Analyse nr. 7			Analyse nr. 8		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Kolonialv.	-1,64164	0,03456	-47,5078	-1,97690	0,03351	-59,0025	-1,55338	0,03454	-44,9790	-1,89594	0,03373	-56,2063
Landhandel	-1,74997	0,03469	-50,4462	-2,11027	0,03330	-63,3749	-1,66171	0,03467	-47,9244	-2,03107	0,03353	-60,5710
Kaffe og te, Öl og mineralvann, Andre n. og n.	-1,75906	0,09502	-18,5121	-2,02331	0,09775	-20,6986	-1,68063	0,09602	-17,5033	-1,94901	0,09916	-19,6549
Bakerv.	-1,70921	0,05849	-29,2210	-1,95931	0,05976	-32,7856	-1,64069	0,05903	-27,7944	-1,89573	0,06058	-31,2955
Melk og delikat.	-1,78168	0,04862	-36,6438	-2,10668	0,04881	-43,1609	-1,70088	0,04894	-34,7539	-2,03383	0,04940	-41,1732
Kjøtt og pølsev.	-1,37999	0,04331	-31,8640	-1,60620	0,04399	-36,5155	-1,29401	0,04350	-29,7498	-1,52026	0,04441	-34,2294
Fisk og vilt	-1,74309	0,04814	-36,2071	-1,99655	0,04888	-40,8447	-1,66901	0,04849	-34,4169	-1,92663	0,04948	-38,9358
Blomster	-1,45709	0,04195	-34,7358	-1,61017	0,04300	-37,4489	-1,39621	0,04227	-33,0270	-1,54883	0,04353	-35,5843
Frukt og grønnsaker	-1,49787	0,06788	-22,0654	-1,71671	0,06971	-24,6255	-1,42694	0,06854	-20,8184	-1,64809	0,07068	-23,3180
Tobakk og sjokolade	-1,68376	0,03511	-47,9590	-2,00438	0,03431	-58,4259	-1,59980	0,03514	-45,5236	-1,92748	0,03458	-55,7413
Vin og spirituosa	-	-	-	-	-	-	-	-	-	-	-	-
Detalj, n. og n.	-1,68323	-	-	-2,01845	-	-	-1,59653	-	-	-1,93933	-	-
Tekstil og bok. 1	-1,88848	0,09964	-18,9526	-2,15499	0,10254	-21,0156	-1,81306	0,10071	-18,0034	-2,08441	0,10404	-20,0353
T. og b. 2	-1,72422	0,04317	-39,9439	-1,98557	0,04357	-45,5735	-1,63893	0,04336	-37,8016	-1,90293	0,04401	-43,2430
Ur, opt., musikk	-1,67563	0,03833	-43,7182	-1,79252	0,03939	-45,5084	-1,60204	0,03851	-41,5978	-1,71358	0,03976	-43,0982
Sykepl. og apot.	-1,21580	0,40182	-3,0257	-1,28093	0,41541	-3,0835	-1,13597	0,40651	-2,7944	-1,19096	0,42172	-2,8240
Bøker, pap., kort.	-1,27681	0,16695	-7,6480	-1,56044	0,17228	-9,0577	-1,18888	0,16882	-7,0422	-1,47624	0,17485	-8,4430
Gull og sølv	-1,92569	0,16656	-11,5616	-2,15635	0,17198	-12,5381	-1,85031	0,16845	-10,9841	-2,08329	0,17456	-11,9342

Tabell 4 (forts.)

Gruppe	Analyse nr. 5			Analyse nr. 6			Analyse nr. 7			Analyse nr. 8		
	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi	Reg.koeff.	Std. av.	t-verdi
Möbler og innb.	-1,76497	0,03779	-47,2296	-2,02293	0,03807	-53,1381	-1,69280	0,03781	-44,7670	-1,93049	0,03832	-50,3737
Jern- og fargev.	-1,65266	0,13046	-12,6676	-1,90016	0,13457	-14,1205	-1,57993	0,13192	-11,9763	-1,83142	0,13657	-13,4101
Biler og rekv.	-2,10616	0,08118	-25,9438	-2,37015	0,08336	-28,4328	-2,02571	0,08199	-24,7078	-2,29343	0,08453	-27,1317
Brensel og smørøljer	-2,10716	0,06965	-30,2544	-2,55220	0,07010	-36,4080	-2,01722	0,07025	-28,7153	-2,47696	0,07105	-34,8609
Detalj ellers	-1,66023	0,08557	-19,4015	-1,89926	0,08803	-21,5761	-1,58506	0,08645	-18,3340	-1,82703	0,08928	-20,4633
Detalj utenom n. og n.	-1,76086	-	-	-1,98106	-	-	-1,67630	-	-	-1,89639	-	-
<u>Variabel</u>												
ln Omsetningen	0,27173	0,00874	31,0732	0,49145	0,00397	123,8378	0,24421	0,00668	28,1233	0,47430	0,00392	120,9917
l : Omsetn. <sup>2</sup>	2,14700	0,13080	16,4140	2,51689	0,13454	18,7078						
ln Bruttoinnt.	0,23106	0,00826	27,9631				0,24478	0,00832	29,4295			
Konstantledd		0,22887			0,14496			0,24869			0,16254	

I kolonne merket "Std. av." står estimatet for standardavviket på regresjonskoeffisienten.

Kolonnen merket "t-verdi" brukes til å teste hypotesen om regresjonskoeffisienten er lik 0 eller ikke. Studentfordelingen har n - s frihetsgrader der n er totalt antall observasjoner, og s er antall variable som inngår i regresjonen.

Tabell nr. 5: Kurveligningen for hver under- og hovedgruppe. Kurven har forme

$$E(Y|X=x) = kx^\gamma e^{\beta x^{-2}} \quad (5.1)$$

der Y er den avhengige variabel og X er omsetningen. Se ellers avsnitt 2. Tallene refererer seg til ligning (2.8).

Analyse nr.	2	4	6	8
$\gamma$	0,95091	0,94000	0,49145	0,47430
$\beta$	1,60084		2,51689	
Gruppe	k	k	k	k
Nærings- og nydelsesmidl. 1	0,1788	0,1935	0,1409	0,1594
N. og n. 2	0,1707	0,1835	0,1289	0,1445
Tekstil og bekl.	0,2523	0,2718	0,1605	0,1804
Ur, opt., musikk	0,2911	0,3131	0,1584	0,1780
Sykepl. og apot.	-	-	-	-
Pap., leketøy og kortev.	0,2726	0,2913	0,1916	0,2126
Gull og sølv, portefølje og kosmetikk	0,3088	0,3312	0,1935	0,2161
Møbler og innbo	0,2308	0,2581	0,1301	0,1552
Jern- og fargev.	0,2530	0,2716	0,1871	0,2070
Bygn.art.	0,2022	0,2182	0,1611	0,1817
Biler og rekv.	0,3053	0,3298	0,1734	0,1958
Mask. og utstyr	0,2576	0,2785	0,1609	0,1819
Div. råvarer	0,2408	0,2569	0,1500	0,1660
Brensel og smøreoljer	0,3100	0,3338	0,2250	0,2528
Skrap, filler	0,4223	0,4656	0,2169	0,2528
Lev. dyr og andre varer	0,2834	0,3024	0,1670	0,1849
Engros	0,250	0,255	0,164	0,185
Kolonialv.	0,1630	0,1735	0,1601	0,1767
Landhandel	0,1462	0,1555	0,1401	0,1544

(forts.)

Tabell nr. 5 (forts.)

Analyse nr.	2	4	6	8
$\gamma$	0,95091	0,94000	0,49145	0,47430
$\beta$	1,60084		2,51689	
Gruppe	k	k	k	k
Kaffe og te, Øl og mineralvann, Andre n. og n.	0,2216	0,2350	0,1528	0,1676
Bakerv.	0,2356	0,2481	0,1629	0,1767
Melk og delikat.	0,1704	0,1805	0,1406	0,1539
Kjøtt og pølsev.	0,2613	0,2791	0,2320	0,2573
Fisk og vilt	0,2322	0,2455	0,1569	0,1713
Blomster	0,3586	0,3770	0,2310	0,2500
Frukt og grønnsaker	0,2697	0,2850	0,2077	0,2264
Tobakk og sjokolade	0,1736	0,1844	0,1557	0,1712
Vin og spirituosa	-	-	-	-
Detalj, n. og n.	0,163	0,173	0,154	0,169
Tekstil og bek. 1	0,2195	0,2321	0,1340	0,1463
T. og b. 2	0,2244	0,2392	0,1587	0,1755
Ur, opt., musikk	0,4193	0,4459	0,1925	0,2120
Sykepl. og apot.	0,5247	0,5618	0,3211	0,3576
Bøker, pap., kort.	0,2038	0,2174	0,2428	0,2688
Gull og sølv	0,2563	0,2715	0,1338	0,1465
Møbler og innb.	0,2483	0,2663	0,1530	0,1707
Jern- og fargev.	0,2382	0,2517	0,1729	0,1885
Biler og rekv.	0,2219	0,2356	0,1080	0,1187
Brensel og smøreoljer	0,1013	0,1075	0,0901	0,0988
Detalj ellers	0,2472	0,2617	0,1730	0,1893
Detalj utenom n. og n.	0,268	0,286	0,159	0,177

## 6. Tabellkommentarer og konklusjoner

Som en vil se er de tabellene som er satt opp høyst ufullstendige i forhold til det som er regnet ut. Grunnen er hovedsakelig tidsnød. Derfor har jeg bare tatt med de tallene som kan ha størst interesse i tabellene.

Et lite teknisk krumspring jeg har vært nødt til å gjøre er å sette logaritmen til alle variable som er lik 0 lik 0. Dette medfører en liten skjevhet i materialet, men må antas å ha liten betydning selvom jeg vet at blant de første 1000 bedrifter er det 8 eller 9 som har hatt omsetning lik 0. Likeledes er  $1 : \text{Omsetningen}^2$  satt lik 0 for omsetningen lik 0.

En større skjevhet gir sannsynligvis det faktum at de 11 384 bedrifter som her er med, ikke er tilfeldig utplukket av totalmassen. Vi må regne med at de bedrifter som først sender inn sine opplysninger er en annen "type" enn de som somler. Hvordan dette gir seg utslag er vanskelig å si, men en indikator har jeg i og med at jeg har gjort en prøvekjøring på de første 1 000 bedrifter i materialet. Her viser det seg at den gjennomsnittlige bruttoinntekt og gjennomsnittlig antall sysselsatte har sunket noe, mens gjennomsnittlig omsetning har steget. En annen side av saken er at blant de første 1 000 bedrifter fantes nesten ikke detaljhandel.

- - -

Vi går så over til de konklusjoner som kan gjøres, og vil da først se på sammenhengen mellom resultatene og Egede Larssens teoretiske utredninger.

Vi skal estimere korrelasjonen mellom de betingede forventninger gitt stratum for totalomsetning og det kjennetegn som skal undersøkes, og vi skal estimere  $C$  definert i (1.1). Den forannevnte korrelasjon defineres som

$$\rho_b = \frac{E(\xi_h - \xi)(\eta_h - \eta)}{\sqrt{E(\xi_h - \xi)^2 E(\eta_h - \eta)^2}} \quad (6.1)$$

Som vi ser må  $\xi_h$ ,  $\xi$ ,  $\eta_h$  og  $\eta$  estimeres, og jeg har som estimatorer brukt de respektive gjennomsnitt innen strata og innen hovedgrupper. Disse blir i stratum nr.  $h$  for bruttoinntekt betegnet med  $\bar{X}_h$ , for antall sysselsatte med  $\bar{Y}_h$  og for omsetningen med  $\bar{Z}_h$ . Gjennomsnittene innen hver hovedgruppe har samme betegnelser, men uten indekser. Vi får nå følgende estimatorer for korrelasjonen mellom de betingede forventninger gitt stratum for totalomsetning og bruttoinntekt:

$$\hat{\rho}_b = \frac{E(\bar{Z}_h - \bar{Z})(\bar{X}_h - \bar{X})}{\sqrt{E(\bar{Z}_h - \bar{Z})^2 E(\bar{X}_h - \bar{X})^2}} \quad (6.2)$$

og for den tilsvarende C:

$$\hat{C} = \sqrt{\frac{E(\bar{Z}_h - \bar{Z})^2}{E(\bar{X}_h - \bar{X})^2}} \frac{\bar{X}}{\bar{Z}} \quad (6.3)$$

der

$$E(\bar{Z}_h - \bar{Z})(\bar{X}_h - \bar{X}) = \sum_h (\bar{Z}_h - \bar{Z})(\bar{X}_h - \bar{X}) \frac{n_h}{n} \quad (6.4)$$

$$E(\bar{Z}_h - \bar{Z})^2 = \sum_h (\bar{Z}_h - \bar{Z})^2 \frac{n_h}{n} \quad (6.5)$$

$$E(\bar{X}_h - \bar{X})^2 = \sum_h (\bar{X}_h - \bar{X})^2 \frac{n_h}{n} \quad (6.6)$$

Vektene  $\frac{n_h}{n}$  er estimatorer for  $\text{Pr}(h)$  i formel (1.2).

Tilsvarende formler for sysselsettingen finnes ved å sette Y-er i stedet for X-er i formlene (6.2-6). Utregningene er satt opp i tabell 6.

Tabell nr. 6

	Engros	Detalj; nærings- og nytelsesmidler	Detalj ellers	I alt
$E(\bar{X}_h - \bar{X})^2$	$2,391 \cdot 10^4$	176,18	376,80	$2,485 \cdot 10^3$
$E(\bar{Y}_h - \bar{Y})^2$	3,49	0,37	0,36	0,67
$E(\bar{Z}_h - \bar{Z})^2$	$2,334 \cdot 10^6$	$5,835 \cdot 10^3$	$17,687 \cdot 10^3$	$23,150 \cdot 10^4$
$E(\bar{X}_h - \bar{X})(\bar{Y}_h - \bar{Y})$	250,19	7,87	11,19	31,56
$E(\bar{X}_h - \bar{X})(\bar{Z}_h - \bar{Z})$	$226,465 \cdot 10^3$	760,13	$2,089 \cdot 10^3$	$22,657 \cdot 10^3$
$E(\bar{Y}_h - \bar{Y})(\bar{Z}_h - \bar{Z})$	$2,171 \cdot 10^3$	34,18	60,83	242,88
$\frac{C_{zx}}{2}$	0,772	0,386	0,718	0,559
$\frac{C_{zy}}{2}$	1,444	0,494	0,849	1,906
$\rho_{zx}$	0,96	0,75	0,81	0,94
$\rho_{zy}$	0,76	0,73	0,76	0,62

Under de forutsetningene som er nevnt i innledningen, kan, med enkelte forbehold som jeg skal komme tilbake til straks, korrigerede gjennomsnitts metode med fordel brukes der hvor }-ene i denne tabellen er større enn de tilsvarende  $\frac{C}{2}$ . Dette gjelder altså ved estimering av bruttoinntekten for alle hovedgrupper, mens for estimering av antall sysselsatte bare for "detalj".

Dessverre blir jeg nå nødt til å ta visse forbehold. Da min undersøkelse bygger på et utvalg som kan være skjevt sammensatt, vil mine estimater være beheftet med feil som vi ikke har kontroll over. Det er likevel ikke sikkert dette har så mye å bety, da korrelasjonen er ganske mye større enn  $\frac{C}{2}$  i alle de tilfellene hvor vi har kommet til konklusjonen at korrigerede gjennomsnitts metode er best. En feil i estimatene for de forskjellige gjennomsnitt må derfor være stor for å forandre på konklusjonen.

Neste forbehold gjelder sammenhengen mellom teori og praksis. Egede Larssen har i sitt notat antatt like mange enheter innen hvert stratum. Dette er som enhver kan se en forutsetning som ikke er oppfylt i vårt utvalg. Jeg har i mine beregninger gitt hvert stratum vekt etter hvor mange enheter det inneholder ( $Pr(h)$  i (1.2) er estimert med  $\frac{n_h}{n}$ ). Dette vil få som konsekvens at strata med mange enheter vil dominere over strata med få enheter. Anta f.eks. at korrigerede gjennomsnitts metode er best å bruke i et stratum som i vår undersøkelse har mange enheter, mens ukorrigerede gjennomsnitts metode er best i et stratum som, fremdeles i vår undersøkelse, har få enheter. Dette vil, hvis vi skal ta konsekvensen av de forannevnte konklusjoner, resultere i at vi bruker korrigerede gjennomsnitts metode også i det siste stratum hvor ukorrigerede gjennomsnitts metode hadde vært best.

Hvilke konsekvenser for denne undersøkelsen ville en oppdeling i omsetningsgrupper ha hatt? Egede Larssen har jo i sitt notat forutsatt denne oppdelingen, men p.g.a. tekniske vansker har jeg måttet droppe dette. De regresjonsanalysene som er foretatt, viser at vi har så godt som lineær regresjon mellom bruttoinntekten og omsetningen over hele variasjonsområdet for omsetningen. Denne linearitet vil nok være enda tydeligere innen en enkelt omsetningsgruppe. Vi må regne med at  $E(\xi_h - \xi)^2$  i (1.1) vil minke betydelig ved en oppdeling i omsetningsgrupper. La oss nå se hva dette vil føre til i C. Vi antar p.g.a. den nevnte linearitet at

$$\eta_h = a_h \xi_h \quad \text{og} \quad \eta = a \xi \quad (6.7)$$



Dette gir for C ved innsetting:

$$C = \left[ E (\xi_h - \xi)^2 \mid E \left( \frac{a_h}{a} \xi_h - \xi \right)^2 \right]^{\frac{1}{2}} \quad (6.8)$$

$\frac{a_h}{a}$  vil her ikke være avhengig av omsetningsgruppens størrelse. Hvis  $E (\xi_h - \xi)^2$  minker, må vi anta at  $E \left( \frac{a_h}{a} \xi_h - \xi \right)^2$  vil minke tilsvarende, og dermed vil C holde seg noenlunde konstant ved en oppdeling i omsetningsgrupper. Ser vi tilbake på  $\rho_b$  i (6.1), vil denne ved samme oppdeling øke, da kovariansen i telleren vil holde seg nokså konstant, mens variansene i nevneren vil minke.

Går vi nå over til å se på sysselsettingen sammenlignet med omsetningen, har vi ikke noen lineær regresjonskurve. Kurven ser ut som øvre halvdel av en parabel gjennom origo med horisontal akse. Vi kan da tilnærmet sette

$$\eta_h = a_h \xi_h^{\frac{1}{2}} \quad \text{og} \quad \eta = a \xi^{\frac{1}{2}} \quad (6.9)$$

hvilket for C gir:

$$C = \left[ E (\xi_h - \xi)^2 \mid E \left( \frac{a_h}{a} \sqrt{\xi_h \xi} - \xi \right)^2 \right]^{\frac{1}{2}} \quad (6.10)$$

Det er klart at ved en oppdeling i omsetningsgrupper, vil den gjennomsnittlige forskjell mellom  $\xi_h$  og  $\xi$  minke.  $\xi_h$  "går mot"  $\xi$  m.a.o., og både teller og nevner vil minke. Derfor er det heller ikke her urimelig å tro at C vil holde seg noenlunde konstant i forhold til  $\rho_b$  som vil øke.

Etter dette skulle vi m.a.o., hvis vi hadde oppdelt materialet i omsetningsgrupper, få enda sterkere grunn til å bruke korrigerede gjennomsnitts metode der hvor vi har funnet at denne er best. Videre er det mulig at vi også hadde funnet denne metoden best der vi nå ikke kan anbefale den.

Jeg vil nå gå over til å behandle de andre resultatene som er kommet frem og begynner med tabell nr. 1. Denne tabellen gir resultatene av tester på om alle regresjonskoeffisientene i en regresjon er lik 0. For stor F-verdi må vi forkaste denne hypotesen, og vi ser av tabellen at i alle regresjonene er det grunn til å forkaste. For eksempel er  $F_{30, \infty} = 2,07$  der  $F_{30, \infty}$  er 99,95 prosent fraktilen i Fisher-fordelingen med 30 frihetsgrader i telleren og uendelig mange frihetsgrader i nevneren. Her må en være oppmerksom på at selv om vi forkaster en hypotese om at alle regresjonskoeffisientene er lik 0, vil ikke dette si at ikke en eller flere kan være lik 0. Jeg har bare ikke testet hypoteser om at bare et utvalg har betydning. Hovedkonklusjonen av tabell nr. 1 er altså at alle regresjonene er rimelige.

Jeg har tidligere vært inne på at antall bedrifter innen de forskjellige næringsgrupper varierer betydelig; noe vi ser av tabell nr. 2. Dette måtte vi selvsagt vente oss, men enkelte grupper har så få observasjoner at de resultater vi har kommet frem til her liten eller ingen mening har. At det i noen grupper ikke finnes observasjoner i det hele tatt er bare å beklage. De gruppene jeg først og fremst vil utelukke er innen engros: 626 Sykepleie og apotekervarer; innen detalj, nærings- og nydelsesmidler: 6428 Vin og spirituosa og innen detalj ellers: 646 Sykepleie- og apotekervarer, 647 Bøker, papir og kortvarer, 648 Gull- og sølvvarer.

Ved første øyekast kan enkelte varianser virke uhyggelig store i tabell 2, men her må vi huske at hele observasjonsområdet for en variabel innen en gruppe er med. Enkelte store avvik øker derfor variansen mye.

Korrelasjonskoeffisientene varierer ganske mye fra gruppe til gruppe. For korrelasjonen mellom antall sysselsatte og bruttoinntekt har vi høyeste verdi for "Papir, leketøy og kortvarer" innen engros med 0,986. Her er forresten korrelasjonskoeffisienten mellom omsetning og bruttoinntekt noe mindre, nemlig 0,982. Dette resultat var kanskje ikke å vente. Lavest verdi har vi for "Brensel og smøreoljer" innen detalj med 0,519. For korrelasjonen mellom bruttoinntekt og omsetning har vi tilsvarende: høyest: 0,998 for "Biler og rekvisita" innen engros og lavest: 0,530 for "Levende dyr og andre varer" innen engros. Denne store variasjon vil jeg komme tilbake til senere.

Tabellene 3 og 4 gir regresjonskoeffisientene og jeg vil her referere til kapital 2 formel (2.7). Regresjonskoeffisientene for hver hovedgruppe trenger en spesiell omtale. Disse er nemlig regnet ut ved et veiet gjennomsnitt av regresjonskoeffisientene for alle de undergrupper som hører med i vedkommende hovedgruppe. Vektene er den relative hyppighet for hver undergruppe innen en hovedgruppe. (F.eks. er vekten til regresjonskoeffisienten for "Nærings- og nydelsesmidler 1" innen engros  $87/1074$ ). Intuitivt er dette riktig nok, men her må vi huske på at vektene er stokastiske variable med ukjent varians.

Tabell nr. 5 gir selve kurveligningene og er regnet ut på grunnlag av tabell 3 og 4. At  $\beta$  mangler i analyse 4 og 8 betyr at jeg her a priori har forutsatt  $\beta = 0$ .  $\beta$  har ellers som man vil se bare vesentlig betydning i de aller laveste omsetningsgrupper, men t-testen viser at den ikke i noe tilfelle helt kan neglisjeres. Parameteren  $\gamma$  gir kurvens krumning hvis  $\beta = 0$ . Med bruttoinntekten som avhengig variabel ser vi at  $\gamma = 0,94$  og dermed at kurven krummer seg veldig lite. (For  $\gamma = 1$  har vi en rett linje). Dette var vel omtrent hva vi kunne vente, og vi har altså så godt som lineær regresjon

mellom bruttoinntekten og omsetningen. Av dette kan vi slutte at våre normalitetsforutsetninger sannsynligvis er riktige. Med antall sysselsatte som avhengig variabel tilsvarende krumningen nesten en liggende parabel ( $f(x) = x^{0,5}$ ). I begge tilfeller er omsetningen den sikre variabel (forklaringsvariabelen). For store omsetninger kan vi også si at  $\gamma$  gir kurvens krumning for  $\beta \neq 0$ .

$k$ -ene er et mål på stigningen av kurvene og viser en del interessante resultater. La oss f.eks. se på analyse nr. 2. De varehandelsgrupper som har størst verdi av  $k$  er de som har størst avanse på varene sine og vice versa de som har minst verdi har også minst avanse. Av de resultater som kan tillegges noen betydning, ser vi at gruppen "skrap og filler" innen engros har størst avanse, men av tabell 1 ser vi også at denne gruppen har minst gjennomsnittsinntekt innen engros. Minst avanse har "Brensel og smøreoljer" innen detalj. Denne gruppen har også minst bruttoinntekt av alle, så klagen bensinfirmaene kommer med er kanskje ikke så helt grunnløse.

De konklusjonene vi kan trekke av analyse 6 og 8 kan virke noe kunstige. Grupper med stor  $k$  som øker omsetningen, må øke antall sysselsatte mer enn grupper med liten  $k$ . Det naturlige spørsmål må jo være: Hvor mange ekstra må man ansette innen en gruppe for å øke omsetningen med en viss verdi? Altså i realiteten den omvendte problemstilling. Dette siste og mest realistiske spørsmål kunne bare vært besvart (selvsagt gjelder dette bare gjennomsnittsverdier innen en gruppe) hvis korrelasjonskoeffisienten mellom antall sysselsatte og omsetningen er lik 1. I grupper med stor korrelasjonskoeffisient kan vi likevel få visse indikatorer.

- - -

Spørsmålet er nå om vi har fått noe særlig utbytte av denne undersøkelsen, og det mener jeg absolutt at vi har, på tross av manglene jeg har nevnt.

P.g.a. utvalgets skjeve sammensetning kan vi vel ikke si at vi har bevist at korrigerede gjennomsnitts metode er best der resultatene har falt ut til fordel for denne metoden, men vi har fått meget sterke indisier på at den er bedre enn vanlige gjennomsnitt i enkelte tilfeller.

På grunnlag av de resultatene jeg har kommet frem til her, kunne det kanskje være en idé å gjennomføre en mer fullstendig undersøkelse med visse modifikasjoner. Vi ser av både tabell 2 og 5 at vi har veldig store variasjoner fra gruppe til gruppe. Å kjøre hver gruppe for seg hadde vært det beste, men for å spare arbeid kunne en kanskje også samle de gruppene som er likest (med likest regresjonskurver) og kjøre dem sammen. Oppdelingen i grupper i min undersøkelse er jo bare konvensjonell og bygger ikke på noen analyse.

Mitt største problem som jeg p.g.a. tidsnød måtte gi opp, var å dele

materialet opp i omsetningsgrupper. Dette bør i en fullstendig undersøkelse gjennomføres, da vi som antydte tidligere, kan oppnå resultater til fordel for korrigerede gjennomsnitts metode der vi ikke har fått dem i denne undersøkelsen.

Bør så selve regresjonsmodellen forandres? Det skulle etter min mening ikke være bryet verd da variansanalysene i tabell nr. 1 har vist berettigelsen av denne modellstrukturen. Videre er det hevet over enhver tvil at leddet  $\beta x^{-2}$  i (2.4) har sin betydning. (Vi kan bare se på t-verdien for  $\beta$ ). At  $\beta$  er blitt så liten i vår analyse, kommer nok av at vi har testet alle varehandelsgruppene simultant. Dette kan også hjelpes på ved å dele opp materialet.

Som konklusjon kan sies at vi i denne analysen har funnet en matematisk modell som tilfredsstillende krav vi setter til statistisk signifikans og som belyser visse relasjoner innen varehandel. Videre har vi funnet meget sterke indisier på at Egede Larssens metode for estimering av visse gjennomsnitt innen varehandel hvis vi har lite utvalg, i enkelte tilfeller er bedre enn rent stratumgjennomsnitt. Noe mer kan en vel ikke forlange å få ut av en såvidt liten undersøkelse som denne.

- - -

#### Litteratur:

E. Sverdrup: Lov og tilfeldighet, bind 1 og 2. Om regresjonsanalyse og minste kvadraters metode.

Harold W. Watts: Statistisk testing av hypoteser ved regresjonsberegninger, Statistisk Sentralbyrås håndbok nr. 22.