

# SOS3003

## Eksamensoppgåver

Oppgave 2 gitt våren 2003

Erling Berge

### **OPPGAVE 2 Logistisk regresjon (teller 50%)**

- Den avhengige variabelen i analysen er innvenn, som fanger opp om en har venner blant innvandrere eller ikke. Dataene er hentet fra en nylig utført norsk intervjuundersøkelse.
- a) Skriv ut ligningen for modellen (ta med forutsetningene).
- b) Beskriv sammenhengen mellom alder og det å ha venner blant innvandrere.
- c) Beskriv sammenhengen mellom ekteskaplig status og det å ha venner blant innvandrere.
- d) Er det samspill mellom kjønn og region i modellen?
- e) Finn oddsratioet for å ha venner blant innvandrere mellom en person på 16 år og en på 70 år.
- f) Finn oddsratioet for å ha venner blant innvandrere mellom menn og kvinner i Oslo og i Trøndelag. Hva forteller de to oddsratioene?
- g) Lag et betinget effektdiagram i logitskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning.
- h) Lag et betinget effektdiagram i sannsynlighetsskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning.

a) Skriv ut ligningen for modellen (ta med forutsetningene) (1) kva er det spurt etter?

Når vi skal formulere ein modell må vi

- definere elementa som inngår i modellen (variablar og datamaterialet)
- definere relasjonane mellom elementa (regresjonslikninga), og
- presisere kva føresetnader som ein må gjere for å bruke modellen

Vi skal skrive ut likninga og føresetnadene

a) Skriv ut ligningen for modellen (ta med forutsetningene) (2) likninga

- I populasjonen føreset vi at det er eit logistisk samband mellom sannsynet for å ha verdien  $Y=1$  på den avhengige variabelen  $Y=INNVEEN$  og dei uavhengige  $X$ -variablane.
- Modellen er da definert ved at vi for kvart case "i" i populasjonen lar  
 $\Pr[Y_i=1 | L_i^*] = E[Y_i]$ , der  $Y_i=1/(1+\exp\{-L_i^*\}) + \varepsilon_i$ ,  $\varepsilon_i$  er feilleddet,  $L_i^*$  er estimert forventa verdi av logiten,  $L_i$ , definert ved
- $L_i^* = E[L_i]=\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_{K-1} X_{(K-1)i}$  der  $K-1$  er talet på uavhengige variable i modellen
- Her er  $Y_i$  og  $X_{1i} - X_{18i}$  definert ved

Variabel symbol	Variabel: navn og tolking, "i" identifiserer eintydig ein respondent i populasjonen
$Y_i$	INNVENN = 1 hvis i har vener mellom innvandrara, 0 elles
$X_{1i}$	MANN = 1 hvis i er mann, 0 elles
$X_{2i}$	ALDER = alder i år for i
$X_{3i}$	UTDAAR = år med fullført utdanning for i
$X_{4i}$	EKSTATUS(1) = 1 hvis i er gift, 0 elles
$X_{5i}$	EKSTATUS(2) = 1 hvis i er skilt, 0 elles
$X_{6i}$	EKSTATUS(3) = 1 hvis i er enke/ enkemann, 0 elles
$X_{7i}$	REGIONNO(1) = 1 hvis i bur i Oslo/ Akershus, 0 elles
$X_{8i}$	REGIONNO(2) = 1 hvis i bur i Hedmark/ Oppland, 0 elles
$X_{9i}$	REGIONNO(3) = 1 hvis i bur i Sydøst-Norge (utanom region 1 og 2), 0 elles
$X_{10i}$	REGIONNO(4) = 1 hvis i bur i Agder/ Rogaland, 0 elles
$X_{11i}$	REGIONNO(5) = 1 hvis i bur i Vest-Norge, 0 elles
$X_{12i}$	REGIONNO(6) = 1 hvis i bur i Trøndelag, 0 elles
$X_{13i}$	REGIONNO(1)*MANN = 1 hvis i er mann og bur i region 1, 0 elles
$X_{14i}$	REGIONNO(2)*MANN = 1 hvis i er mann og bur i region 2, 0 elles
$X_{15i}$	REGIONNO(3)*MANN = 1 hvis i er mann og bur i region 3, 0 elles
$X_{16i}$	REGIONNO(4)*MANN = 1 hvis i er mann og bur i region 4, 0 elles
$X_{17i}$	REGIONNO(5)*MANN = 1 hvis i er mann og bur i region 5, 0 elles
$X_{18i}$	REGIONNO(6)*MANN = 1 hvis i er mann og bur i region 6, 0 elles

Vår 2004 © Erling Berge 5

a) Skriv ut ligningen for modellen (ta med forutsetningene) (3) estimat av likninga

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1						
MANN	-.440	.307	2.055	1	.152	.644
ALDER	-.034	.004	74.333	1	.000	.966
UTDAAR	.119	.016	54.847	1	.000	1.126
EKSTATUS			8.066	3	.045	
EKSTATUS(1)	.353	.131	7.294	1	.007	1.424
EKSTATUS(2)	.383	.194	3.888	1	.049	1.467
EKSTATUS(3)	.287	.271	1.124	1	.289	1.332
REGIONNO			18.549	6	.005	
REGIONNO(1)	-.058	.277	.045	1	.833	.943
REGIONNO(2)	-.159	.325	.239	1	.625	.853
REGIONNO(3)	-.427	.279	2.352	1	.125	.652
REGIONNO(4)	-.821	.287	8.189	1	.004	.440
REGIONNO(5)	-.180	.280	.412	1	.521	.835
REGIONNO(6)	-.841	.331	6.471	1	.011	.431
REGIONNO * MANN			22.173	6	.001	
REGIONNO(1) by MANN	.228	.378	.366	1	.545	1.257
REGIONNO(2) by MANN	-.266	.446	.355	1	.551	.767
REGIONNO(3) by MANN	.421	.379	1.234	1	.267	1.524
REGIONNO(4) by MANN	1.047	.393	7.089	1	.008	2.848
REGIONNO(5) by MANN	-.356	.378	.889	1	.346	.700
REGIONNO(6) by MANN	.479	.436	1.209	1	.272	1.615
Constant	1.459	.280	27.182	1	.000	4.300

Vår 2004 © Erling Berge 6

a) Skriv ut ligningen for modellen (ta med forutsetningene)  
(4) estimat av likninga

- $L_i^* = 1.459 - 0.440MANN_i - 0.034ALDER_i + 0.119UTDAAR_i + 0.353EKSTATUS(1)_i + 0.383EKSTATUS(2)_i + 0.287EKSTATUS(3)_i - 0.058REGIONNO(1)_i - 0.159REGIONNO(2)_i - 0.427REGIONNO(3)_i - 0.821REGIONNO(4)_i - 0.180REGIONNO(5)_i - 0.841REGIONNO(6)_i + 0.228REGIONNO(1)_i * MANN_i - 0.266REGIONNO(2)_i * MANN_i + 0.421REGIONNO(3)_i * MANN_i + 1.047REGIONNO(4)_i * MANN_i - 0.356REGIONNO(5)_i * MANN_i + 0.479REGIONNO(6)_i * MANN_i$
- Residualen,  $e_i$ , finn vi som  $e_i = Y_i - 1/(1+\exp\{-L_i^*\})$

a) Skriv ut ligningen for modellen (ta med forutsetningene) (5) føresetnadene

Ein føreset at

1. modellen er rett spesifisert, dvs.:
  1. den funksjonelle forma for alle betinga sannsyn for  $Y=1$  er logistiske funksjonar av  $X$ -ane (dette svarar til at Logiten er lineær i parametrane)
  2. ingen relevante variablar er utelatne
  3. ingen irrelevante variablar er inkluderte
2. alle  $X$ -variablane er utan målefeil
3. alle case er uavhengige
  - det er ikkje perfekt multikollinearitet
  - det er ikkje perfekt diskriminering
  - stort nok utval

## b) Beskriv sammenhengen mellom alder og det å ha venner blant innvandrere.

- $Y_i = 1$  tyder at personen i har "noen eller mange" vener mellom innvandarane
- Logitkoeffesienten for alder er  $-0.034$  og signifikant ulik 0
- Direkte tyder dette at logiten minkar med  $0.034$  logiteiningar for kvart år eldre personen er. Det negative forteiknet tyder at sannsynet vil minke med aukande alder
- For å seie meir om korleis sannsynet for at  $Y=1$  endrar seg må vi sjå på oddsraten når personar vert eitt år eldre, OR (sannsynet for  $Y=1$  ved 1 års auke i alder) =  $\exp\{-0.034\} = 0.966$
- Av dette ser vi at sannsynet minkar med  $3,4\%$  ved 1 års auke i alder
- Dersom vi ser på skilnaden mellom 20 åringar og 60 åringar vert oddsraten  $\exp\{-0.034 \cdot 40\} = 0.256$
- Endringa i prosent er da  $(1 - \exp\{-0.034 \cdot 40\}) \cdot 100 = 74,4\%$

## c) Beskriv sammenhengen mellom ekteskaplig status ogt det å ha venner blant innvandrere.

- Ekteskapeleg status har ein signifikant verknad i modellen
- Referanse kategorien for ekteskapeleg status er alle personar som seier dei er ugift. Koeffesientane for EKSTATUS(j) fortel om gruppa j skil seg frå referanse kategorien ugift
- Med 5% testnivå ser vi at enke/ enkemenn ( $j=3$ ) ikkje har ein koeffesient ulik 0. Denne gruppa skil seg da ikkje frå referansegruppa i sannsyn for å vener mellom innvandarar. Både gifte ( $j=1$ ) og skilte ( $j=2$ ) har koeffesientar som er signifikant ulik 0
- Koeffesientane for både gifte og skilte er positive slik at sannsynet for å ha vener mellom innvandarane er større enn i referansegruppa
- Oddsraten mellom gifte og dei som ikkje er gifte er  $\exp\{0.353\} = 1.424$  og viser at gifte har  $42\%$  høgare sannsyn for å hav innvandar vener enn ikkje gifte om alt anna er likt
- Oddsraten mellom skilte og dei som ikkje er skilte er  $\exp\{0.383\} = 1.467$  og viser at skilte har  $47\%$  høgare sannsyn for å hav innvandar vener enn ikkje skilte om alt anna er likt

## d) Er det samspill mellom kjønn og region i modellen?

- Samspelsledda har samla ein signifikant verknad i modellen. Referansekategorien er kvinner busett i Nord-Norge
- Berre eitt av samspelsledda har ein effekt som er signifikant ulik referansekategorien

e) Finn oddsratioet for å ha venner blant innvandrere mellom en person på 16 år og en på 70 år

- Oddsraten for å ha vener mellom innvandrarar mellom 16 åringar og 70 åringar kan finnast som oddsraten  $\exp\{-0.034*(70-16)\} = \exp\{-0.034*54\} = \exp\{-1,836\} = 0,159$

f) Finn oddsratioet for å ha venner blant innvandrere mellom menn og kvinner i Oslo og i Trøndelag. Hva forteller de to oddsratioene? (1) Oslo

- Odds for å ha venter mellom innvandrere for menn i Oslo =  $\exp\{L(\text{menn i Oslo})\}$  der  $L(\text{menn i Oslo}) = 1.459 - 0.440\text{MANN}_i - 0.034\text{ALDER}_i + 0.119\text{UTDAAR}_i + 0.353\text{EKSTATUS}(1)_i + 0.383\text{EKSTATUS}(2)_i + 0.287\text{EKSTATUS}(3)_i - 0.058\text{REGIONNO}(1)_i - 0.159\text{REGIONNO}(2)_i - 0.427\text{REGIONNO}(3)_i - 0.821\text{REGIONNO}(4)_i - 0.180\text{REGIONNO}(5)_i - 0.841\text{REGIONNO}(6)_i + 0.228\text{REGIONNO}(1)_i * \text{MANN}_i - 0.266\text{REGIONNO}(2)_i * \text{MANN}_i + 0.421\text{REGIONNO}(3)_i * \text{MANN}_i + 1.047\text{REGIONNO}(4)_i * \text{MANN}_i - 0.356\text{REGIONNO}(5)_i * \text{MANN}_i + 0.479\text{REGIONNO}(6)_i * \text{MANN}_i$
- $L(\text{menn i Oslo}) = 1.459 - 0.440 - 0.034\text{ALDER}_i + 0.119\text{UTDAAR}_i + 0.353\text{EKSTATUS}(1)_i + 0.383\text{EKSTATUS}(2)_i + 0.287\text{EKSTATUS}(3)_i - 0.058 + 0.228$
- Tilsvarende for kvinner i Oslo
- $L(\text{kvinner i Oslo}) = 1.459 - 0.034\text{ALDER}_i + 0.119\text{UTDAAR}_i + 0.353\text{EKSTATUS}(1)_i + 0.383\text{EKSTATUS}(2)_i + 0.287\text{EKSTATUS}(3)_i - 0.058$

f) Finn oddsratioet for å ha venner blant innvandrere mellom menn og kvinner i Oslo og i Trøndelag. Hva forteller de to oddsratioene? (2) Oslo og Trøndelag

- Oddsraten:  $\frac{\exp\{L(\text{menn i Oslo})\}}{\exp\{L(\text{kvinner i Oslo})\}} = \frac{\exp\{L(\text{menn i Oslo}) - L(\text{kvinner i Oslo})\}}{\exp\{L(\text{menn i Oslo}) - L(\text{kvinner i Oslo})\}} = \exp\{1.459 - 0.440 - 0.034\text{ALDER}_i + 0.119\text{UTDAAR}_i + 0.353\text{EKSTATUS}(1)_i + 0.383\text{EKSTATUS}(2)_i + 0.287\text{EKSTATUS}(3)_i - 0.058 + 0.228 - (1.459 - 0.034\text{ALDER}_i + 0.119\text{UTDAAR}_i + 0.353\text{EKSTATUS}(1)_i + 0.383\text{EKSTATUS}(2)_i + 0.287\text{EKSTATUS}(3)_i - 0.058)\} = \exp\{-0.440 + 0.228\} = \exp\{-0,212\} = 0,81$
- Tilsvarende for menn og kvinner i Trøndelag:
- Oddsraten =  $\exp\{-0.440 + 0.479\} = \exp\{0.039\} = 1,04$
- Mangelen på signifikans for interaksjonledd og variabelen Mann kan truleg tilskrives multikollineariteten i modellen

g) Lag et betinget effektdiagram i logitskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning

koeffesientar Region	konstant	Region	Region Mann	Gifte	Mann	Gifte Menn	Ikkje Gift Mann	Gift Kvinne	Ikkje Gift Kvinne
Oslo og Akershus	1,459	-0,058	0,228	0,353	-0,44	1,542	1,189	1,754	1,401
Hedmark og Oppland	1,459	-0,159	-0,266	0,353	-0,44	0,947	0,594	1,653	1,300
SørøstNorge	1,459	-0,427	0,421	0,353	-0,44	1,366	1,013	1,385	1,032
Agder og Rogaland	1,459	-0,821	1,047	0,353	-0,44	1,598	1,245	0,991	0,638
VestNorge	1,459	-0,180	-0,356	0,353	-0,44	0,836	0,483	1,632	1,279
Trøndelag	1,459	-0,841	0,479	0,353	-0,44	1,010	0,657	0,971	0,618
Nord-Norge	1,459	0	0	0,353	-0,44	1,372	1,019	1,812	1,459

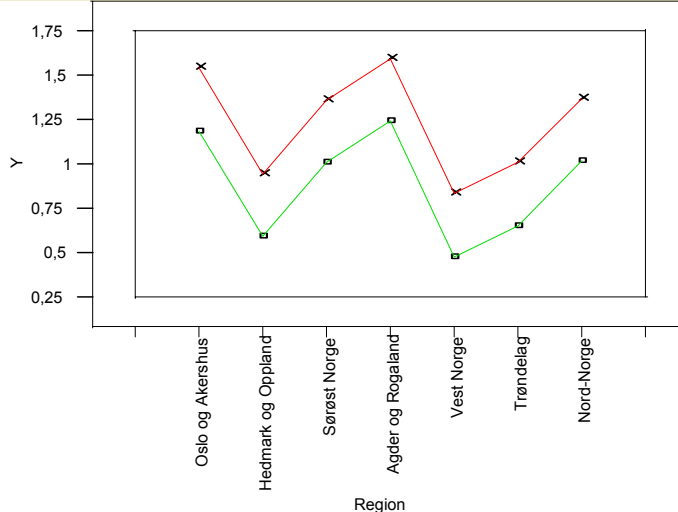
Vår 2004

© Erling Berge

15

Overlay Plot

Overlay Y's



Y x— Gifte Menn

■— Ikkje Gift Mann

Vår 2004

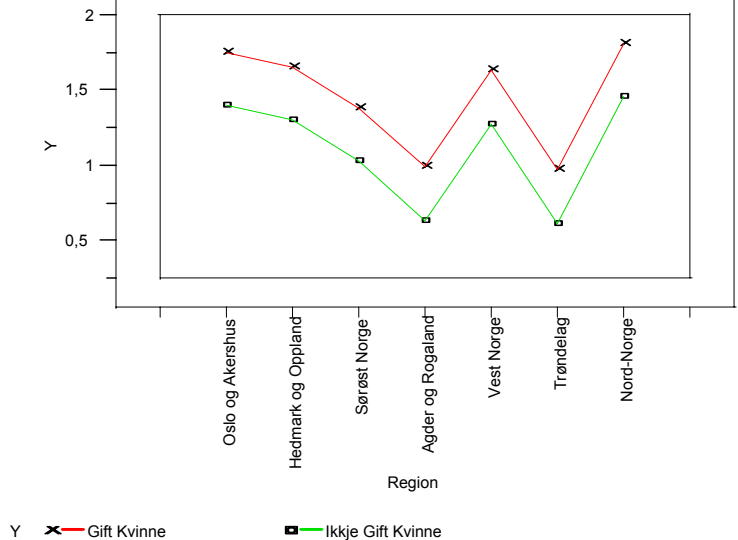
© Erling Berge

16



Overlay Plot

Overlay Y's



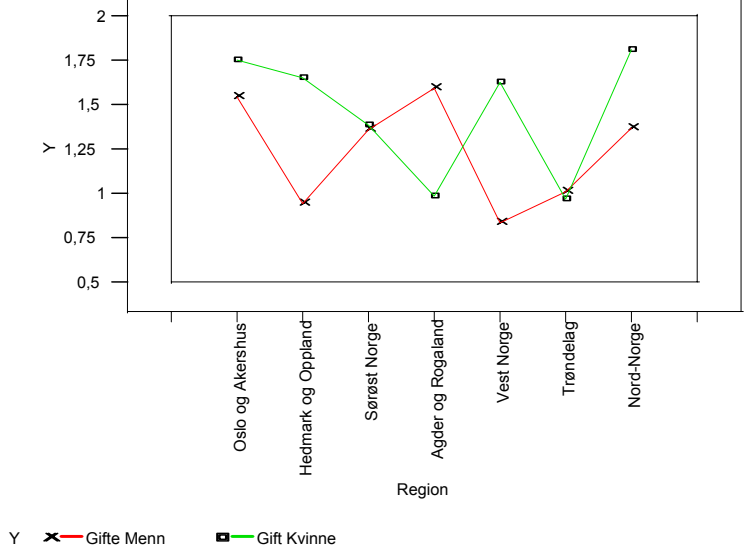
Vår 2004

© Erling Berge

17

Overlay Plot

Overlay Y's



Vår 2004

© Erling Berge

18

h) Lag et betinget effektdiagram i sannsynlighetsskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning (1)

- EKSTATUS(1) = 1, alle andre EKSTATUS(1) er 0
- Gjennomsnittlig alder er 45,79 år
- Gjennomsnittleg utdanning er 4,4 år
- Set vi dette inn i den estimerte logiten,  $L_i^* = 1.459 - 0.440MANN_i - 0.034*45.79 + 0.119*4.4 + 0.353*1 - 0.058REGIONNO(1)_i - 0.159REGIONNO(2)_i - 0.427REGIONNO(3)_i - 0.821REGIONNO(4)_i - 0.180REGIONNO(5)_i - 0.841REGIONNO(6)_i + 0.228REGIONNO(1)_i * MANN_i - 0.266REGIONNO(2)_i * MANN_i + 0.421REGIONNO(3)_i * MANN_i + 1.047REGIONNO(4)_i * MANN_i - 0.356REGIONNO(5)_i * MANN_i + 0.479REGIONNO(6)_i * MANN_i = 0.77 + 0.440MANN_i - 0.058REGIONNO(1)_i - 0.159REGIONNO(2)_i - 0.427REGIONNO(3)_i - 0.821REGIONNO(4)_i - 0.180REGIONNO(5)_i - 0.841REGIONNO(6)_i + 0.228REGIONNO(1)_i * MANN_i - 0.266REGIONNO(2)_i * MANN_i + 0.421REGIONNO(3)_i * MANN_i + 1.047REGIONNO(4)_i * MANN_i - 0.356REGIONNO(5)_i * MANN_i + 0.479REGIONNO(6)_i * MANN_i$

h) Lag et betinget effektdiagram i sannsynlighetsskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning (2)

- For gifte menn blir logiten  $L_{GM} = 0.77 - 0.44 + (0.228 - 0.058)REGIONNO(1)_i - (0.266 + 0.159)REGIONNO(2)_i + (0.421 - 0.427)REGIONNO(3)_i + (1.047 - 0.821)REGIONNO(4)_i - (0.356 + 0.180)REGIONNO(5)_i + (0.479 - 0.841)REGIONNO(6)_i$
- For gifte kvinner blir logiten  $L_{GK} = 0.77 - 0.058REGIONNO(1)_i - 0.159REGIONNO(2)_i - 0.427REGIONNO(3)_i - 0.821REGIONNO(4)_i - 0.180REGIONNO(5)_i - 0.841REGIONNO(6)_i$

h) Lag et betinget effektdiagram i sannsynlighetsskalaen som viser regionale forskjeller i det å ha venner blant innvandrere for gifte personer med gjennomsnittlig alder og utdanning (3)

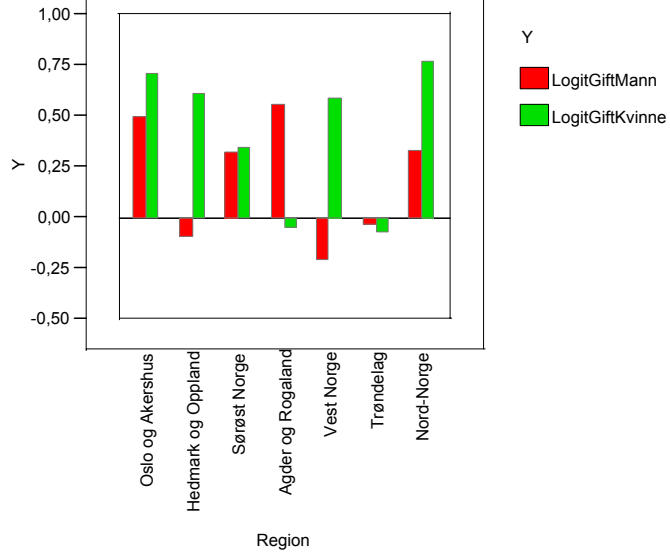
- $L_{GM} = 0.33 + 0.17\text{REGIONNO}(1)_i - 0.425\text{REGIONNO}(2)_i - 0.007\text{REGIONNO}(3)_i + 0.226\text{REGIONNO}(4)_i - 0.536\text{REGIONNO}(5)_i - 0.362\text{REGIONNO}(6)_i$
- For gifte kvinner blir logiten  $L_{GK} = 0.77 - 0.058\text{REGIONNO}(1)_i - 0.159\text{REGIONNO}(2)_i - 0.427\text{REGIONNO}(3)_i - 0.821\text{REGIONNO}(4)_i - 0.180\text{REGIONNO}(5)_i - 0.841\text{REGIONNO}(6)_i$

$$\text{Sannsyn} = 1/(1+\exp\{-L\})$$

Region(j)=1	Konst menn	Konst kvinner	region	Region *menn	Logit gifte menn	Logit gifte kvinner	Sannsyn gifte menn	Sannsyn gifte kvinner
Oslo og Akershus	0,33	0,77	-0,058	0,228	0,5	0,712	0,622459	0,670843
Hedmark og Oppland	0,33	0,77	-0,159	-0,266	-0,095	0,611	0,476268	0,648169
SørøstNorge	0,33	0,77	-0,427	0,421	0,324	0,343	0,580299	0,584919
Agder og Rogaland	0,33	0,77	-0,821	1,047	0,556	-0,051	0,635527	0,487252
VestNorge	0,33	0,77	-0,180	-0,356	-0,206	0,59	0,448681	0,643365
Trøndelag	0,33	0,77	-0,841	0,479	-0,032	-0,071	0,492001	0,482258
Nord-Norge	0,33	0,77	0	0	0,33	0,77	0,581759	0,683521

Chart

Overlay Chart



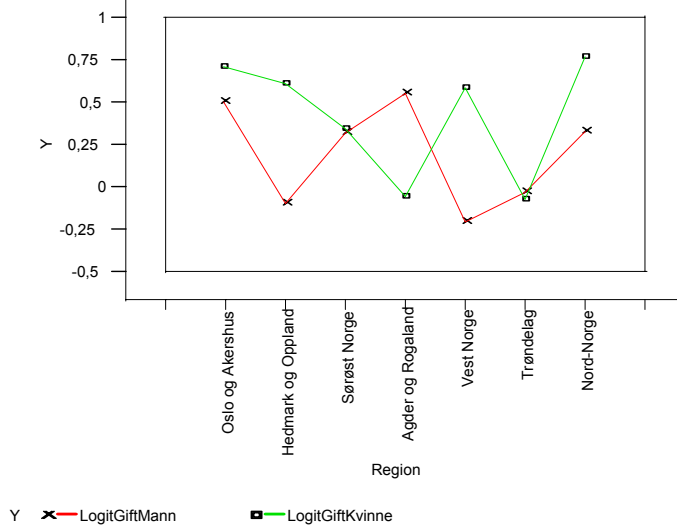
Vår 2004

© Erling Berge

23

Overlay Plot

Overlay Y's



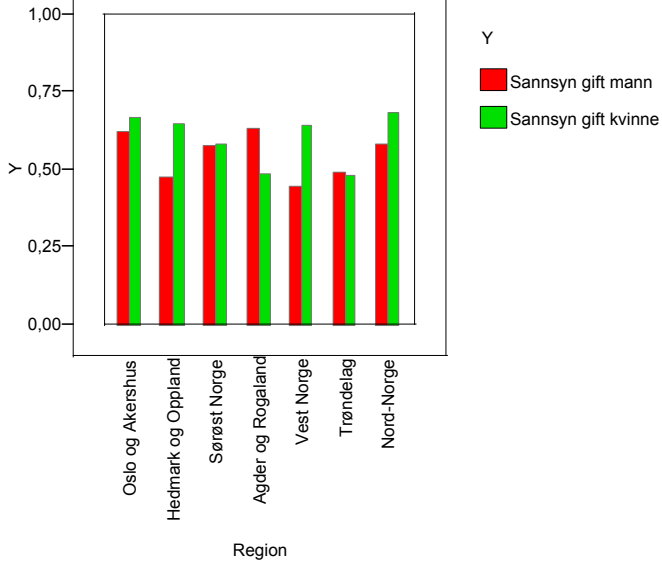
Vår 2004

© Erling Berge

24

Chart

Overlay Chart



Overlay Plot

Overlay Y's

