

SOS3003

Eksamensoppgåver

Gjennomgang våren 2004

Erling Berge

Gjennomgang av

- Oppgåve 2 gitt hausten 2003

Haust 2003 Oppg ve 2

Den avhengige variabelen i den logistiske regresjonsanalysen er *freegl*, som har verdien 1 hvis respondenten var helt enig i utsagnet "Gay men and lesbians should be free to live their own life as they wish." Ellers har variabelen verdien null.

- a) Skriv ut ligningen for modellen (ta med forutsetningene).
- b) Beskriv sammenhengen mellom alder og holdning til homofile og lesbiske.
- c) Beskriv sammenhengen mellom ekteskaplig status og holdning til homofile og lesbiske.
- d) Er det samspill mellom land og kj nn i modellen?
- e) Finn oddsratioene for   være helt enig at homofile og lesbiske b r f  leve som de vil mellom kvinner og menn i Storbritannia, Norge og Polen. Hva forteller de tre oddsratioene?
- f) Finn predikert sannsynlighet for   være helt enig i at homofile og lesbiske b r f  leve som de vil for en 50  r gammel ugift mann bosatt i Norge.

Haust 2003 2a Ligningen for modellen (med forutsetningene)

FREEGL Gay and lesbians free to live as they wish

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	.00 Agree,Not agree	4989	80.5	83.3	83.3
	1.00 Agree strongly	997	16.1	16.7	100.0
	Total	5986	96.6	100.0	
Missing	System	212	3.4		
Total		6198	100.0		

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
FEMALE	6198	.00	1.00	.5011	.50004
AGE	6183	15.00	98.00	45.7189	18.19806
MARRIED Married (1), not married (0)	6185	.00	1.00	.5243	.49945
CNTRY3	6198	1.00	3.00	2.0094	.81947
Valid N (listwise)	6174				

Haust 2003 2a Ligningen for modellen (med forutsetningene)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	FEMALE	-.076	.158	.233	1	.629	.926
	AGE	-.024	.002	112.660	1	.000	.976
	MARRIED	-.176	.076	5.397	1	.020	.838
	CNTRY3			29.259	2	.000	
	CNTRY3(1)	.549	.147	13.987	1	.000	1.732
	CNTRY3(2)	.741	.138	28.948	1	.000	2.098
	CNTRY3 * FEMALE			16.588	2	.000	
	CNTRY3(1) by FEMALE	.602	.199	9.102	1	.003	1.826
	CNTRY3(2) by FEMALE	.778	.193	16.329	1	.000	2.178
	Constant	-1.202	.134	80.600	1	.000	.301

a. Variable(s) entered on step 1: FEMALE, AGE, MARRIED, CNTRY3, CNTRY3 * FEMALE .

Haust 2003 2a Ligningen for modellen (med forutsetningene)

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

Haust 2003 2a likninga (1)

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

Haust 2003 2a likninga (2)

- Modellen er da definert ved at vi lar
Observert verdi av $Y_i =$ predikert verdi av Y_i
+ residualen $=1/(1+\exp\{-L_i^*\}) + e_i$,
 e_i er residualen, L_i^* er estimert forventa
verdi av logiten, L_i , definert ved
- $L_i^* = -1.202 - 0.076 * \text{Female} - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * \text{Female} + 0.778 * \text{NO} * \text{Female}$

Haust 2003 2a estimatet av logiten

Y= 1 if agree strongly	B	S.E.	Wald	df	Sig.	Exp(B)
FEMALE	-.076	.158	.233	1	.629	.926
AGE	-.024	.002	112.660	1	.000	.976
MARRIED	-.176	.076	5.397	1	.020	.838
CNTRY3			29.259	2	.000	
CNTRY3(1) = GB	.549	.147	13.987	1	.000	1.732
CNTRY3(2) = NO	.741	.138	28.948	1	.000	2.098
CNTRY3 * FEMALE			16.588	2	.000	
CNTRY3(1) by FEMALE	.602	.199	9.102	1	.003	1.826
CNTRY3(2) by FEMALE	.778	.193	16.329	1	.000	2.178
Constant	-1.202	.134	80.600	1	.000	.301

Haust 2003 2a føresetnader

Ein føreset at

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

Haust 2003 2a (ekstrastoff): elementa i modellen

- Populasjonen vi granskar er befolkningane i dei tre landa Storbritannia (GB), Norge (NO) og Polen (PO)
- Det skal lagast ein modell av den avhengige variabelen $Y = \text{freegl}$, som har verdien 1 dersom respondenten var heilt einig i utsegna "Gay men and lesbians should be free to live their own life as they wish.", elles har variabelen verdien null
- Det vil bli argumentert med at følgjande uavhengige variablar verkar inn på korleis ein svarar på spørsmålet:

Haust 2003 2a (ekstrastoff): elementa i modellen

Variabel symbol	Variabel: namn og tolking, "i" identifiserer eintydig ein respondent i populasjonen
X_{1i}	FEMALE = 1 dersom respondenten i er kvinne, 0 elles
X_{2i}	AGE = alder i år for respondenten i
X_{3i}	MARRIED = 1 dersom respondenten i er gift, 0 elles
X_{4i}	CNTRY3 = 1 dersom respondenten i er frå Great Britain, 2 dersom respondenten i er frå Norge, og 3 dersom respondenten i er frå Polen

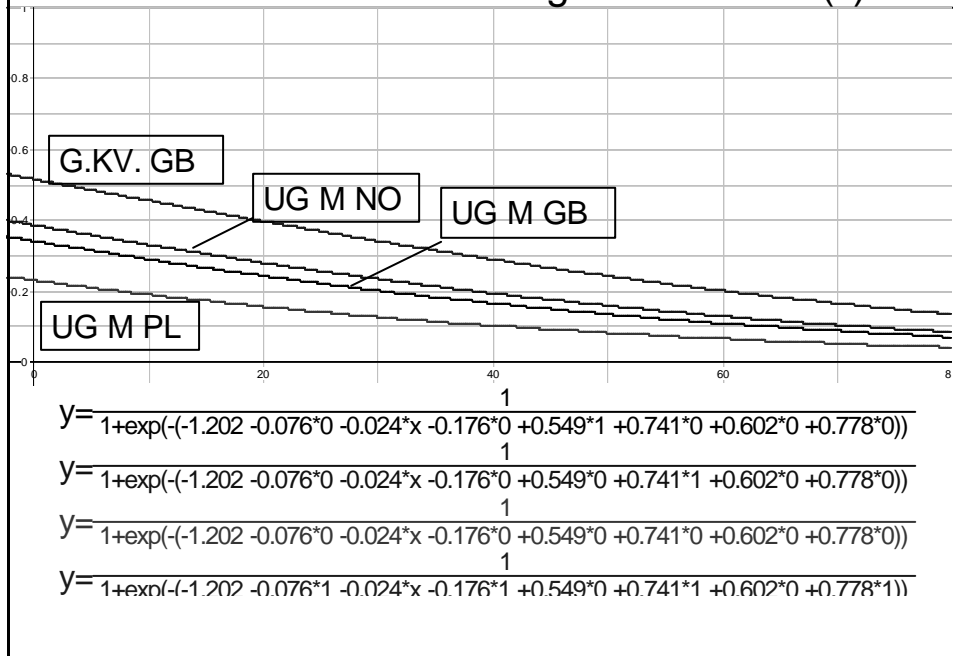
Haust 2003 2b samanhengen med alder

- Koeffesienten i logiten er negativ: -0.024
dvs sannsynet minkar med aukande alder alt
anna likt
- Oddsen for $Y=1$ minkar med
 $100(\exp[-0.024]-1) \%$
for kvart år eldre personen er
- Ein 70åring vil ha $100(\exp[-0.024(70-20)]-1) \%$
mindre odds enn ein 20 åring
- $\exp(-0.024*50)=\exp(-1.2) = 0.301$

Haust 2003 2b samanhengen med alder: sannsynet

- Predikert verdi av sannsynet for at $Y_i=1$:
 $\Pr[Y_i=1 | L_i^*] = 1/(1+\exp\{-L_i^*\})$
der $L_i^* = -1.202 -0.076*Female -0.024*Age$
 $-0.176*Married +0.549*GB +0.741*NO$
 $+0.602*GB*Female +0.778*NO*Female$
- Dersom vi set inn $Female = 0$, $Married = 0$
vil alderssamanhengen i dei ulike landa sjå
ut som på neste side (siste plottet gjeld
gifte kvinner i Storbritannia)

Haust 2003 2b samanhengen med alder(x)



Haust 2003 2 c) samanhengen med ekteskapeleg status

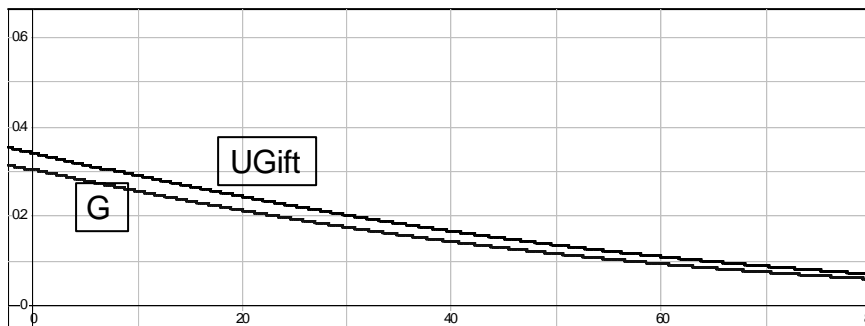
- Koeffesienten for ekteskapeleg status er -0.176. Det tyder at sannsynet vil minke når ein går frå verdien 0 til verdien 1 på variabelen ekteskapeleg status. Gifte er i mindre grad einige i utsegna "Gay men and lesbians should be free to live their own life as they wish"
- Oddsraten mellom ugifte og gifte er =.838 slik at oddsen for Y=1 er $100(\exp[-0.176]-1) \% = 16.2\%$ mindre for gifte enn for ugifte

Haust 2003 2 c) sammenhengen med ekteskapeleg status: sannsynet

- Predikert verdi av sannsynet for at $Y_i = 1$:

$$\Pr[Y_i = 1 \mid L_i^*] = 1 / (1 + \exp\{-L_i^*\})$$
 der $L_i^* = -1.202 - 0.076 * \text{Female} - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * \text{Female} + 0.778 * \text{NO} * \text{Female}$
- Dersom vi for $\text{GB} = 1$ set inn $\text{Female} = 0$,
 og lar ekteskapeleg status og alder variere vil sammenhengen sjå ut som på neste side

Haust 2003 2 c) sammenhengen med ekteskapeleg status for menn i GB etter alder og ekteskapeleg status

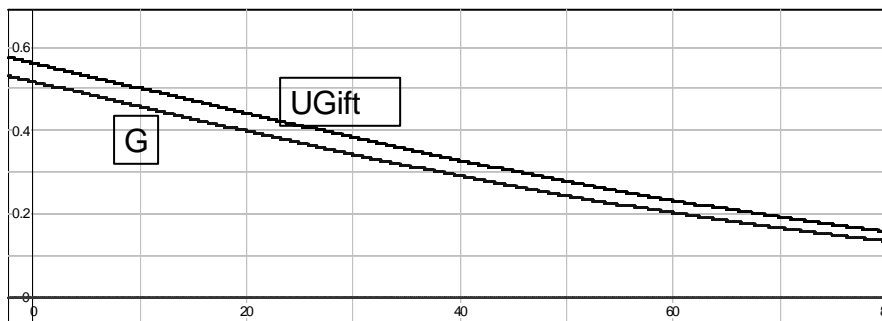


$$y = \frac{1}{1 + \exp(-(-1.202 - 0.076 * 0 - 0.024 * x - 0.176 * 0 + 0.549 * 1 + 0.741 * 0 + 0.602 * 0 + 0.778 * 0))}$$

$$y = \frac{1}{1 + \exp(-(-1.202 - 0.076 * 0 - 0.024 * x - 0.176 * 1 + 0.549 * 1 + 0.741 * 0 + 0.602 * 0 + 0.778 * 0))}$$

$y = 0$

Haust 2003 2 c) sammenhengen med ekteskaplig status i kvinner i NO etter alder og ekteskapeleg status



$$Y = \frac{1}{1 + \exp(-(-1.202 - 0.076 * 1 - 0.024 * x - 0.176 * 0 + 0.549 * 0 + 0.741 * 1 + 0.602 * 0 + 0.778 * 1))}$$

$$Y = \frac{1}{1 + \exp(-(-1.202 - 0.076 * 1 - 0.024 * x - 0.176 * 1 + 0.549 * 0 + 0.741 * 1 + 0.602 * 0 + 0.778 * 1))}$$

y=0

Haust 2003 2d estimatet av logiten

Y= 1 if agree strongly	B	S.E.	Wald	df	Sig.	Exp(B)
FEMALE	-.076	.158	.233	1	.629	.926
AGE	-.024	.002	112.660	1	.000	.976
MARRIED	-.176	.076	5.397	1	.020	.838
CNTRY3			29.259	2	.000	
CNTRY3(1) = GB	.549	.147	13.987	1	.000	1.732
CNTRY3(2) = NO	.741	.138	28.948	1	.000	2.098
CNTRY3 * FEMALE			16.588	2	.000	
CNTRY3(1) by FEMALE	.602	.199	9.102	1	.003	1.826
CNTRY3(2) by FEMALE	.778	.193	16.329	1	.000	2.178
Constant	-1.202	.134	80.600	1	.000	.301

Haust 2003 2d) Er det samspel mellom land og kjønn i modellen?

- Samspelet involverer 2 variablar som må testast under eitt
- I tabellen som gir estimat av logiten finn vi ei linje der det står "CNTRY3 * FEMALE". Lenger ut på linja finn vi oppgitt at Wald-observatoren har verdien 16.588
- Wald-observatoren vil her gi oss ein test av samspelet. Dei to samspelsledda vert testa under eitt av SPSS
- Wald-observatoren er Kji-kvadratfordelt med talet av fridomsgrader lik talet av variablar i testen (i dette høvet 2: sjå kolonnen for df)

Haust 2003 2d) Er det samspel mellom land og kjønn i modellen? Testen

- Dersom nullhypotesen om at vi ikkje har interaksjon er rett, er sannsynet for å finne ein Wald-observator på 16.588 mindre enn 0.000 i følgje tabellen ovanfor (sjå kolonnen for "sig.")
- Vi forkastar nullhypotesen og vil inntil vidare meine at det er samspel mellom land og kjønn i modellen

Haust 2003 2e)

Oddsratene mellom kvinner og menn

- e) Finn oddsratene for å vere helt einig i at homofile og lesbiske bør få leve som dei vil mellom kvinner og menn i Storbritannia, Norge og Polen. Kva fortel dei tre oddsratene?
- Oddsraten er raten mellom to sannsyn
- Vi har i denne modellen
 $\Pr[Y_i = 1 | L_i^*] = 1/(1 + \exp\{-L_i^*\})$
der $L_i^* = -1.202 - 0.076 * \text{Female} - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * \text{Female} + 0.778 * \text{NO} * \text{Female}$

Haust 2003 2e) Oddsraten mellom kvinner og menn

Oddsraten mellom to verdiar av L, t.d. L* og L**

- Sidan oddsen er definert som

$$O_i [Y_i = 1 | L_i^*] = \frac{\Pr[Y_i = 1 | L_i^*]}{1 - \Pr[Y_i = 1 | L_i^*]} = \exp(L_i^*)$$

er **ODDSRATEN**

$$\frac{O_i [Y_i = 1 | L_i^*]}{O_i [Y_i = 1 | L_i^{**}]} = \frac{\exp(L_i^*)}{\exp(L_i^{**})} = \exp(L_i^* - L_i^{**})$$

Haust 2003 2e) Oddsraten mellom kvinner og menn

- Estimert av logiten er

$$L_i = -1.202 - 0.076 * \text{Female} - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * \text{Female} + 0.778 * \text{NO} * \text{Female}$$

- For kvinner finn vi logiten

$$L_i^* = -1.202 - 0.076 * 1 - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * 1 + 0.778 * \text{NO} * 1 =$$

$$-1.278 - 0.024 * \text{Age} - 0.176 * \text{Married} + (0.549 + 0.602) * \text{GB} + (0.741 + 0.778) * \text{NO} =$$

Haust 2003 2e) Oddsater mellom kvinner og menn

- For menn finn vi logiten

$$L_i^{**} = -1.202 - 0.076 * 0 - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * 0 + 0.778 * \text{NO} * 0 =$$

$$-1.202 - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO}$$

- Da er differansen mellom kvinner og menn

$$L_i^* - L_i^{**} = -1.278 - 0.024 * \text{Age} - 0.176 * \text{Married} + (0.549 + 0.602) * \text{GB} + (0.741 + 0.778) * \text{NO} - (-1.202 - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO}) =$$

$$-1.278 - 0.024 * \text{Age} - 0.176 * \text{Married} + (0.549 + 0.602) * \text{GB} + (0.741 + 0.778) * \text{NO} + 1.202 + 0.024 * \text{Age} + 0.176 * \text{Married} - 0.549 * \text{GB} - 0.741 * \text{NO} =$$

$$-1.278 + 1.202 - 0.024 * \text{Age} + 0.024 * \text{Age} - 0.176 * \text{Married} + 0.176 * \text{Married} + (0.549 + 0.602 - 0.549) * \text{GB} + (0.741 + 0.778 - 0.741) * \text{NO} =$$

Haust 2003 2e) Oddsraten mellom kvinner og menn

- $L_i^* - L_i^{**} = -0.076 + 0.602 \cdot GB + 0.778 \cdot NO$
- Av dette ser vi at oddsraten mellom kvinner og menn er
- I Storbritannia: **exp[0.526] = 1,692** (kvinner er 69% over menn)
- I Norge: **exp[0.702] = 2,018** (kvinner er 102% over menn)
- I Polen: **exp[-0.076] = 0.926** (odds for kvinner og menn like)
-0.076 er koeffesienten for FEMALE og gir oss effekten av kjønn i Polen når GB=NO=0. Denne koeffesienten er ikkje signifikant ulik 0. Oddsraten mellom kvinner og menn i Polen vil derfor vere tilnærma lik 1
- I utrekningane for Norge og Storbritannia må ein likevel ta omsyne til koeffesienten til FEMALE sidan den påverkar estimatet av alle koeffesientane

Haust 2003 2f) predikert sannsyn for å vere
einig

- f) Finn predikert sannsynlighet for å være helt enig i at homofile og lesbiske bør få leve som de vil for en 50 år gammel ugift mann bosatt i Norge.

Haust 2003 2f) predikert sannsyn for å vere einig 50 årig ugift Norsk mann

Predikert verdi av sannsynet for å vere enig=

- $\Pr[Y_i = 1 | L_i^*] = 1/(1 + \exp\{-L_i^*\})$
der $L_i^* = -1.202 - 0.076 * \text{Female} - 0.024 * \text{Age} - 0.176 * \text{Married} + 0.549 * \text{GB} + 0.741 * \text{NO} + 0.602 * \text{GB} * \text{Female} + 0.778 * \text{NO} * \text{Female}$
- Vi set inn for 50 årig ugift Norsk mann
- $L_i^* = -1.202 - 0.076 * 0 - 0.024 * 50 - 0.176 * 0 + 0.549 * 0 + 0.741 * 1 + 0.602 * 0 * 0 + 0.778 * 1 * 0 = -1.202 - 0.024 * 50 + 0.741 = -1.661$
- $\Pr[Y_i = 1 | L_i^*] = 1/(1 + \exp\{-(-1.661)\}) = 0.16$