

SOS3003

Examination questions

Spring 2004

Erling Berge

Question 2 Introduction

- As part of the study of how those who have victims of crime in the family view the legal system, also the impact of having a victim of crime in the family on the experience of insecurity during walks alone in the dark was studied. The relationship was studied in a multivariate approach controlling for the impact of other factors by means of logistic regression.
- The dependent variable is "Feeling very unsafe walking alone after dark" The variable is coded 1 for those who answer "very unsafe" on the question of "Feeling of safety of walking alone in local area after dark". Those who give other answers are coded 0. Listwise deletion is used for missing data. Eight control variables are introduced. Some results from this analysis are included in appendix tables for Question 2.

Question 2 (Logistic regression, weight 0,5)

- a) Discuss the relation between having “victims of crime in the family” and “Feeling very unsafe walking alone after dark” as expressed by this regression analysis
- b) Find a confidence interval for the regression coefficient of having “victims of crime in the family” with level of significance of 0,01. Test if employment status contributes significantly to the model
- c) Write up the equations for the conditional effect plots of the relationship between age and probability of “feeling very unsafe walking alone after dark” for Spanish and Norwegian women that have experienced crime in the family, have lived for 10 years in the area, have 12 years of education, are employed for wages, and live in a city without partner
- d) Write up the model estimated
- e) Discuss to what degree the assumptions of the logistic regression model has been satisfied
- f) Discuss if there are cases with unreasonably large influence on the regression results

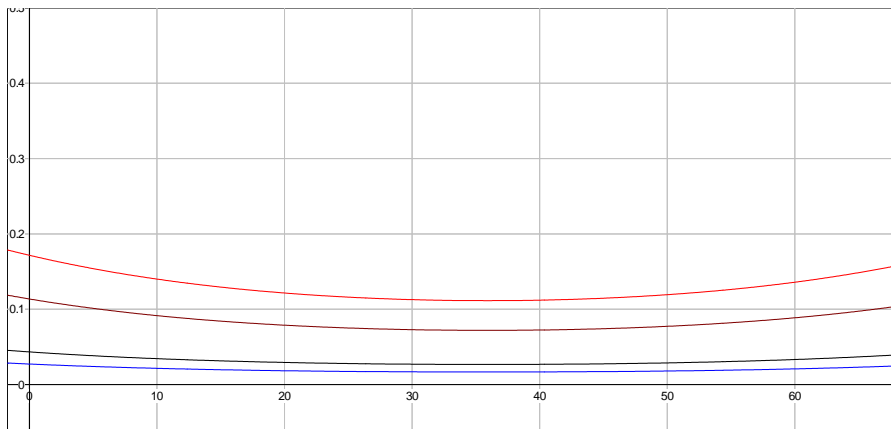
2 a) Discuss the relation between having “victims of crime in the family” and “Feeling very unsafe walking alone after dark” as expressed by this regression analysis

- See table next page
- One might conclude for example from the data presented that crime do not only affect the quality of life for the victim but also to a significant degree the quality of life for family members of the victim. This holds independently of the country and type of locality where one lives, how long one has lived there, gender, age, if one lives with a partner, amount of education, and employment status.
- The result seems robust in relation to alternative explanations such as catholic/ protestant culture (Spain vs other countries), urban/ rural location, single, woman, old, and knowledgeable

Test blocks	2LogLikelihood of Block	Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Block 0	-3556,912	Constant	-,233	,534	,190	1	,663	,792
Block 1	-3545,014	victim	,480	,112	18,275	1	,000	1,616
Block 2	-3194,117	yrlydae	-,004	,003	1,518	1	,218	,996
		eduys	-,165	,027	36,991	1	,000	,848
		female	1,519	,126	144,972	1	,000	4,570
		liveWithPartner	-,159	,111	2,046	1	,153	,853
Block 3	-3150,129	selfempl	,336	,256	1,713	1	,191	1,399
		notempl	,430	,131	10,804	1	,001	1,537
Block 4	-3129,757	age	-,028	,016	3,071	1	,080	,973
		age2	,000	,000	6,388	1	,011	1,000
Block 5	-3029,321	suburb	-,525	,168	9,747	1	,002	,592
		town	-,622	,150	17,172	1	,000	,537
		village	-1,554	,183	72,302	1	,000	,211
		countryside	-1,455	,311	21,892	1	,000	,234
Block 6	-2870,363	Spain	-2,274	,371	37,582	1	,000	,103
		Sweden	-1,722	,510	11,426	1	,001	,179
		Norway	-1,874	,624	9,010	1	,003	,154
Block 7	-2849,161	eduInSpain	,131	,032	16,837	1	,000	1,140
		eduInSweden	,021	,045	,208	1	,648	1,021
		eduInNorway	,017	,052	,107	1	,743	1,017

2 a)

- $\exp(0,48) = 1,616$ shows that having “victims of crime in the family” increase the odds of “Feeling very unsafe walking alone after dark” by 61.6 per cent
- Exploring the relationship by conditional effect plots:



$y=1/(1+\exp(-(-0.233+0.48 \times 1-0.004 \times 10-0.165 \times 20+1.519 \times 0-0.159 \times 0-0.028 \times x+0.00039 \times x \times x)))$	Victim=1; Female=0
$y=1/(1+\exp(-(-0.233+0.48 \times 0-0.004 \times 10-0.165 \times 20+1.519 \times 0-0.159 \times 0-0.028 \times x+0.00039 \times x \times x)))$	Victim=0; Female=0
$y=1/(1+\exp(-(-0.233+0.48 \times 1-0.004 \times 10-0.165 \times 20+1.519 \times 1-0.159 \times 0-0.028 \times x+0.00039 \times x \times x)))$	Victim=1; Female=1
$y=1/(1+\exp(-(-0.233+0.48 \times 0-0.004 \times 10-0.165 \times 20+1.519 \times 1-0.159 \times 0-0.028 \times x+0.00039 \times x \times x)))$	Victim=0; Female=1

2 b) Find a confidence interval for the regression coefficient of having “victims of crime in the family” with level of significance of 0,01. Test if employment status contributes significantly to the model

- Confidence interval
- $b_{\text{victim}} - SEb_{\text{victim}} * t_{1\%} < \beta_{\text{victim}} < b_{\text{victim}} + SEb_{\text{victim}} * t_{1\%}$
- $0,48 - 0,112 * 2,576 < \beta_{\text{victim}} < 0,48 + 0,112 * 2,576$
- $0,48 - 0,288512 < \beta_{\text{victim}} < 0,48 + 0,288512$
- $0,191488 < \beta_{\text{victim}} < 0,768512$

Employment status

- $\chi^2_H = -2\{\log_e \mathcal{L}_{K-H} - \log_e \mathcal{L}_K\}$
- $\chi^2_2 = -2\{\log_e \mathcal{L}_6 - \log_e \mathcal{L}_8\} = -(2\log_e \mathcal{L}_6) + (2\log_e \mathcal{L}_8) = -(-3194,117) + (-3150,129) = 3194,117 - 3150,129 = 43,988$

2 c) Find the formulas for conditional effect plots of the relationship between age and probability of feeling very insecure walking alone after dark for Spanish and Norwegian women that have victims of crime in their family, has lived in the area for 10 years, have 12 years of education, are gainfully employed and live in a city without partner.

- $$L_i = -0,233 + 0,480 \cdot \text{victim}_i - 0,004 \cdot \text{yrlvdae}_i - 0,165 \cdot \text{eduyrs}_i + 1,519 \cdot \text{female}_i - 0,159 \cdot \text{liveWithPartner}_i + 0,336 \cdot \text{selfempl}_i + 0,430 \cdot \text{notempl}_i - 0,028 \cdot \text{age}_i + 0,0004 \cdot \text{age2}_i - 0,525 \cdot \text{suburb}_i - 0,622 \cdot \text{town}_i - 1,554 \cdot \text{village}_i - 1,455 \cdot \text{countryside}_i - 2,274 \cdot \text{Spain}_i - 1,722 \cdot \text{Sweden}_i - 1,874 \cdot \text{Norway}_i + 0,131 \cdot \text{eduInSpain}_i + 0,021 \cdot \text{eduInSweden}_i + 0,017 \cdot \text{eduInNorway}_i$$

2 c) Variable values

	Spain	Norway
victim	1	1
yrlvdae	10	10
eduyrs	12	12
female	1	1
liveWithPartner	0	0
selfempl	0	0
notempl	0	0
age	-	-
age2	-	-
suburb	0	0
town	0	0
village	0	0
countryside	0	0
Spain	1	0
Sweden	0	0
Norway	0	1
eduInSpain	12*1	12*0
eduInSweden	12*0	12*0
eduInNorway	12*0	12*1

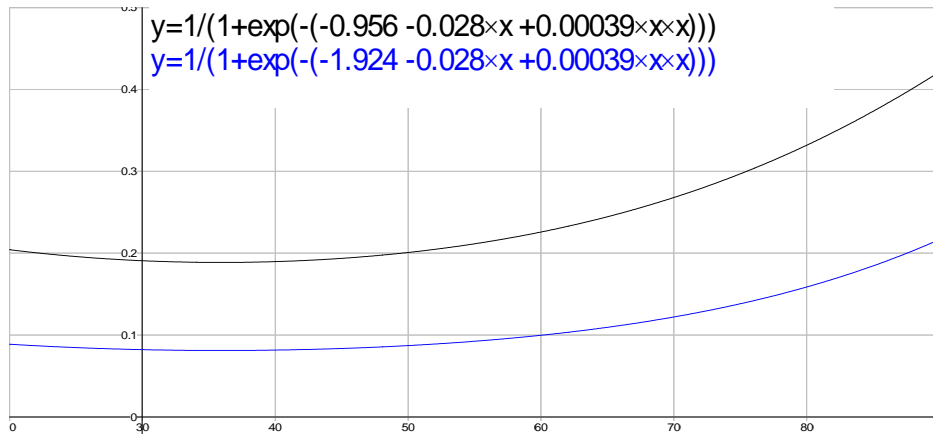
2 c) Conditional effect plots

- $L_i = -0,233 + 0,480 \cdot 1 - 0,004 \cdot 10 - 0,165 \cdot 12 + 1,519 \cdot 1 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2 - 2,274 \cdot \text{Spain}_i - 1,874 \cdot \text{Norway}_i + 0,131 \cdot 12 \cdot \text{Spain}_i + 0,017 \cdot 12 \cdot \text{Norway}_i =$
- $-0,233 + 0,480 - 0,04 - 1,98 + 1,519 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2 - 2,274 \cdot \text{Spain}_i - 1,874 \cdot \text{Norway}_i + 1,572 \cdot \text{Spain}_i + 0,204 \cdot \text{Norway}_i =$
- $-0,254 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2 - 2,274 \cdot \text{Spain}_i - 1,874 \cdot \text{Norway}_i + 1,572 \cdot \text{Spain}_i + 0,204 \cdot \text{Norway}_i$
- $-0,254 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2 - 0,702 \cdot \text{Spain}_i - 1,67 \cdot \text{Norway}_i$

2 c) Conditional effect plots

- For Norwegian women
- $\Pr(Y=1) = 1/(1+\exp\{-(-1,924 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2)\})$
- For Spanish women
- $\Pr(Y=1) = 1/(1+\exp\{-(-0,956 - 0,028 \cdot \text{age}_i + 0,00039 \cdot \text{age}_i^2)\})$

2 c) Conditional effect plots



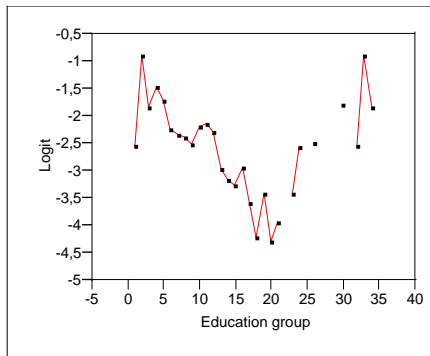
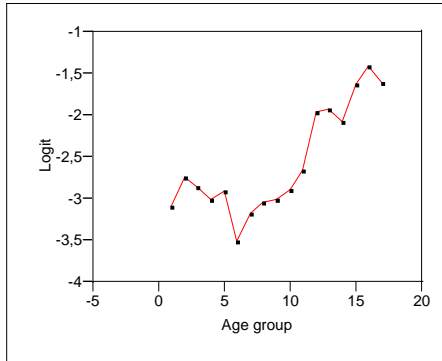
2 d) Write up the model estimated

- See slide 12 above
- $\Pr[Y_i=1] = E[Y_i]$, where $Y_i = [1 / (1 + \exp\{-L_i^*\})] + e_i$

$$L_i^* = E[L_i] = \beta_0 + \sum_{k=1}^{19} \beta_k X_{ki}$$

2 e) Discuss to what degree the assumptions of the logistic regression model has been satisfied

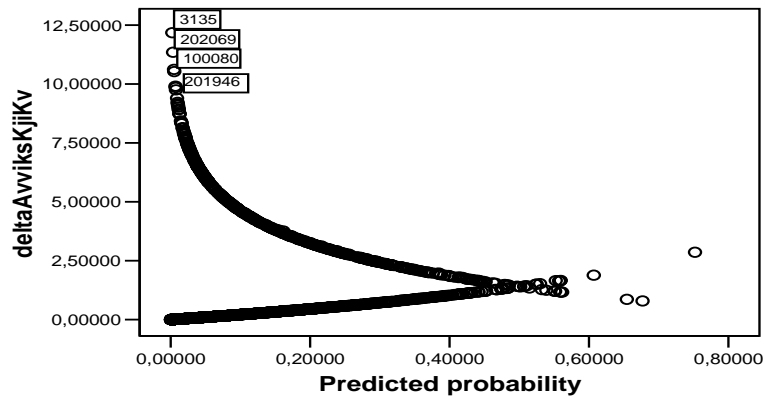
- Linearity of Logit?



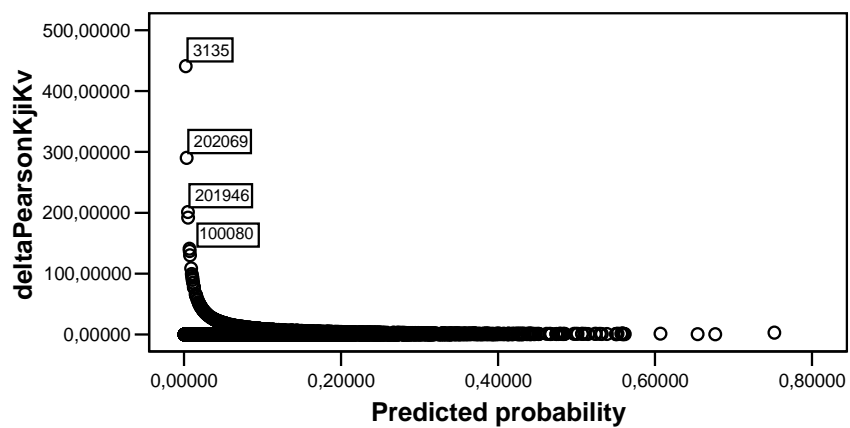
2 e) Irrelevant variables?

Test blocks	2LogLikelihood of Block	Variable	B	S.E.	Wald	df	Sig.	Exp(B)
Block 0	-3556,912	Constant	-.233	,534	,190	1	,663	,792
Block 1	-3545,014	Victim	,480	,112	18,275	1	,000	1,616
Block 2	-3194,117	yr1vdae	-.004	,003	1,518	1	,218	,996
		edyrs	-.165	,027	36,991	1	,000	,848
		female	1,519	,126	144,972	1	,000	4,570
		liveWithPartner	-.159	,111	2,046	1	,153	,853
Block 3	-3150,129	selfempl	,336	,256	1,713	1	,191	1,399
		notempl	,430	,131	10,804	1	,001	1,537
Block 4	-3129,757	Age	-.028	,016	3,071	1	,080	,973
		age2	,000	,000	6,388	1	,011	1,000
Block 5	-3029,321	suburb	-.525	,168	9,747	1	,002	,592
		Town	-.622	,150	17,172	1	,000	,537
		village	-1,554	,183	72,302	1	,000	,211
		countryside	-1,455	,311	21,892	1	,000	,234
Block 6	-2870,363	Spain	-2,274	,371	37,582	1	,000	,103
		Sweden	-1,722	,510	11,426	1	,001	,179
		Norway	-1,874	,624	9,010	1	,003	,154
Block 7	-2849,161	eduInSpain	,131	,032	16,837	1	,000	1,140
		eduInSweden	,021	,045	,208	1	,648	1,021
		eduInNorway	,017	,052	,107	1	,743	1,017

2 f) Discuss if there are cases with unreasonably large influence on the regression results



2 f) Influence



2 f) Influence

IDNO	PRE	COO	deltaPearson KjiKv	deltaAvviks KjiKv
3135	0,0023	0,0621	440,915	12,184
110106	0,0211	0,2372	46,715	7,760
202069	0,0034	0,0613	290,272	11,350
202626	0,0116	0,2527	85,211	8,934