SOS3003
Applied data analysis for social science
Lecture note 08-2010

Erling Berge
Department of sociology and political science
NTNU

Literature

- Logistic regression II
  Hamilton Ch 7 p217-242

Definitions I

- The probability that person no i shall have the value 1 on the variable Y_i will be written $Pr(Y_i = 1)$.
- Then $Pr(Y_i \neq 1) = 1 - Pr(Y_i = 1)$
- The odds that person no i shall have the value 1 on the variable Y_i, here called O_i, is the ratio between two probabilities

$$O_i\left(y_i = 1\right) = \frac{\Pr\left(y_i = 1\right)}{1 - \Pr\left(y_i = 1\right)} = \frac{p_i}{1 - p_i}$$

Definitions II

- The LOGIT , L_i , for person no i (corresponding to $Pr(Y_i=1)$) is the natural logarithm of the odds, O_i , that person no i has the value 1 on variable Y_i . This is written:
  $L_i = \ln(O_i) = \ln\{p_i/(1-p_i)\}$
- The model assumes that L_i is a linear function of the explanatory variables x_{ji} ,
- i.e.:
- $L_i = \beta_0 + \Sigma_j \beta_j x_{ji}$, where j=1,..,K-1, and i=1,..,n

Logistic regression: assumptions

- The model is correctly specified
  - The logit is linear in its parameters
  - All relevant variables are included
  - No irrelevant variables are included
- x-variables are measured without error
- Observations are independent
- No perfect multicollinearity
- No perfect discrimination
- Sufficiently large sample

Assumptions that cannot be tested

- Model specification
  - All relevant variables are included
- x-variables are measured without error
- Observations are independent

Two will be tested automatically.

If the model can be estimated there is

- No perfect multicollinearity and
- No perfect discrimination

LOGISTIC REGRESSION
Statistical problems may be due to

- Too small a sample
- High degree of **multicollinearity**
  - Leading to large standard errors (imprecise estimates)
  - Multicollinearity is discovered and treated in the same way as in OLS regression
- High degree of **discrimination** (or separation)
  - Leading to large standard errors (imprecise estimates)
  - Will be discovered automatically by SPSS

Assumptions that can be tested

- Model specification
  - logit is linear in the parameters
  - no irrelevant variables are included
- Sufficiently large sample
  - What constitutes a sufficiently large sample is not always clear.
  - It depends on how the cases are distributed between 0 and 1 categories. If one of these is too small there will be problems estimating partial effects.
  - It also depends on the number of different patterns in the sample and how cases are distributed across these

LOGISTIC REGRESSION: TESTING (1)

- Testing implies an assessment of whether statistical problems leads to departure from the assumptions

Two tests are useful

- (1) The **Likelihood ratio test** statistic: χ^2_H
  - Can be used analogous to the F-test
- (2) Wald test
  - The square root of this can be used analogous to the t-test

LOGISTIC REGRESSION: TESTING (2)

- The Likelihood Ratio test :
- A ratio between two Likelihoods is equivalent to a difference between two **LogLikelihoods**
- The difference between the **LogLikelihood** ($\mathcal{LL}$) of two **nested** models, estimated on **the same data**, can be used to test which of two models fits the data best, just like the F-statistic is used in OLS regression
- The test can also be used for singe regression coefficients (single variables). In small samples it has better properties than the Wald statistic

LOGISTIC REGRESSION: TESTING (3)

The Likelihood Ratio test statistic

- $\chi^2_H = -2[\mathcal{LL}(\text{model1}) - \mathcal{LL}(\text{model2})]$

will, if the null hypothesis of no difference between the two models is correct, be distributed approximately (for large n) as the chi-square distribution with number of degrees of freedom equal to the difference in number of parameters in the two models (H)

Example of a Likelihood Ratio test

- Model 1: just constant
- Model 2: constant plus one variable

- $\chi^2_H = -2[\mathcal{LL}(\text{model1}) - \mathcal{LL}(\text{model2})]$
  $= -2\mathcal{LL}(\text{model1}) + 2\mathcal{LL}(\text{model2})$
- Find the value of the ChiSquare and the number of degrees of freedom
- e.g.: LogLikelihood (mod1) = 209,212/(-2)
- LogLikelihood (mod2) = 195,267/(-2)

From Tab 7.1: **-2 Log-Likelihood**
209,212
195,684
195,269
195,267
195,267

Logistic Regression in SPSS I

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	153	100.0
	Missing Cases	0	.0
	Total	153	100.0
Unselected Cases		0	.0
Total		153	100.0

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
OPEN	0
CLOSE	1

Logistic Regression in SPSS IIa

Iteration History[a,b,c]

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	209.212	-.275
	2	209.212	-.276
	3	209.212	-.276

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 209.212

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.

Logistic Regression in SPSS IIb

Classification Table[a,b]

Observed			Predicted: SCHOOLS SHOULD CLOSE OPEN	CLOSE	Percentage Correct
Step 0	SCHOOLS SHOULD CLOSE	OPEN	87	0	100.0
		CLOSE	66	0	.0
	Overall Percentage				56.9

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.276	.163	2.864	1	.091	.759

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	lived	12.683	1	.000
	Overall Statistics		12.683	1	.000

Logistic Regression in SPSS IIIa

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients: Constant	lived
Step 1	1	195.684	.376	-.034
	2	195.269	.455	-.041
	3	195.267	.460	-.041
	4	195.267	.460	-.041

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 209.212

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Logistic Regression in SPSS IIIb

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	13.944	1	.000
	Block	13.944	1	.000
	Model	13.944	1	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	195.267[a]	.087	.117

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Logistic Regression in SPSS IIIc

Classification Table[a]

			Predicted		
			SCHOOLS SHOULD CLOSE		Percentage Correct
	Observed		OPEN	CLOSE	
Step 1	SCHOOLS SHOULD CLOSE	OPEN	59	28	67.8
		CLOSE	29	37	56.1
	Overall Percentage				62.7

a. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	lived	-.041	.012	11.399	1	.001	.960
	Constant	.460	.263	3.069	1	.080	1.584

a. Variable(s) entered on step 1: lived.

LOGISTIC REGRESSION: TESTING (4)

The Wald test

- The Wald (or chisquare) test statistic provided by SPSS = $t^2 = (b_k / SE(b_k))^2$ (where t is the t used by Hamilton) can be used for testing single parameters similarly to the t-statistic of the OLS regression
  - Check out Hamilton table 7.1: find t^2, see Wald above
- If the null hypothesis is correct, t will (for large n) in logistic regression be approximately normally distributed
- If the null hypothesis is correct, the Wald statistic will (for large n) in logistic regression be approximately chisquare distributed with df=1

Excerpt from Hamilton Table 7.2

Iterasjon	**-2 Log likelihood**					
0	209,212					
1	152,534					
2	149,466					
3	149,382					
4	149,382					
5	149,382					
Variables	**B**	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
Lived	-,046	,015	9,698	1	,002	,955
Educ	-,166	,090	3,404	1	,065	,847
Contam	1,208	,465	6,739	1	,009	3,347
Hsc	2,173	,464	21,919	1	,000	8,784
Constant	1,731	1,302	1,768	1	,184	5,649

Confidence interval for parameter estimates

- Can be constructed based on the fact that the square root of the Wald statistic approximately follows a normal distribution with 1 degree of freedom
- $b_k - t_\alpha * SE(b_k) < \beta_k < b_k + t_\alpha * SE(b_k)$
  where t_α is a value taken from the table of the **normal distribution** with level of significance equal to α

Can be constructed based on the t-distribution (1)

- If a table of the normal distribution is missing one may use the **t-distribution** since the t-distribution is approximately normally distributed for large n-K (e.g. for n-K > 120)

Excerpt from Hamilton Table 7.3 (from SPSS)

STATA SPSS		B	S.E.	t^2 Wald	df	Prob>t Sig.	Exp(B)
Step 1	lived	-,047	,017	7,550	1	,006	,954
	educ	-,206	,093	4,887	1	,027	,814
	contam	1,282	,481	7,094	1	,008	3,604
	hsc	2,418	,510	22,508	1	,000	11,223
	female	-,052	,557	,009	1	,926	,950
	kids	-,671	,566	1,406	1	,236	,511
	nodad	-2,226	,999	4,964	1	,026	,108
	Constant	2,894	1,603	3,259	1	,071	18,060

More from Hamilton Table 7.3

Iteration		**-2LogLikelihood**	**Coefficients**							
			Const	**lived**	**educ**	**contam**	**hsc**	**female**	**kids**	**nodad**
Step0		**209,212**	**-0,276**							
Step1	**1**	**147,028**	**1,565**	**-,027**	**-,130**	**,782**	**1,764**	**-,015**	**-,365**	**-1,074**
	2	**141,482**	**2,538**	**-,041**	**-,187**	**1,147**	**2,239**	**-,037**	**-,580**	**-1,844**
	3	**141,054**	**2,859**	**-,046**	**-,204**	**1,269**	**2,401**	**-,050**	**-,662**	**-2,184**
	4	**141,049**	**2,893**	**-,047**	**-,206**	**1,282**	**2,418**	**-,052**	**-,671**	**-2,225**
	5	**141,049**	**2,894**	**-,047**	**-,206**	**1,282**	**2,418**	**-,052**	**-,671**	**-2,226**

Is the model in table 7.3 better than the model in table 7.2 ?

- **$\mathcal{LL}$(model in 7.3) = 141,049/(-2)**
- **$\mathcal{LL}$(model in 7.2) = 149,382/(-2)**

- χ^2_H = -2[$\mathcal{LL}$(model 7.2) - $\mathcal{LL}$(model 7.3)]
- Find χ^2_H value
- Find H
- Look up the table of the chisquare distribution

The model of the probability of observing y=1 for person i

$$\Pr(y_i = 1) = E\left[y_i \mid x\right] = \frac{1}{1+\exp\left(-L_i\right)} = \frac{\exp(L_i)}{1+\exp(L_i)}$$

where the logit $L_i = \beta_0 + \sum_{j=1}^{K-1} \beta_j X_{ji}$ is a linear function of the explanatory variables

It is not easy to interpret the meaning of the β coefficients just based on this formula

The odds ratio

- The **odds ratio**, **O**, can be interpreted as the relative effect of having one variable value rather than another
- e.g. if x_{ki} = t+1 in L_i' and x_{ki} = t in L_i
- **O** = O_i (Y_i=1| L_i')/ O_i (Y_i=1| L_i)
  = exp[L_i']/ exp[L_i]
  = exp[β_k]
- Why β_k ?

The odds ratio : example I

- The Odds for answering yes =
  $e^{b_0 + b_1 * Alder + b_2 * Kvinne + b_3 * E.utd + b_4 * Barn\ i\ HH}$
- The odds ratio for answering yes between women and men =

$$\frac{e^{b_0 + b_1 * Alder + b_2 * 1 + b_3 * E.utd + b_4 * Barn\_i\_HH}}{e^{b_0 + b_1 * Alder + b_2 * 0 + b_3 * E.utd + b_4 * Barn\_i\_HH}} = e^{b_2}$$

Remember the rules of power exponents

Conditional Effect Plot

- Set all x-variables except x_k to fixed values and enter these into the equation for the logit
- Plot Pr(Y=1) as a function of x_k i.e.
- P =1/(1+exp[-L]) = 1/(1+exp[-konst - $b_k x_k$])

  for all reasonable values of x_k ,

  “konst” is the constant obtained by entering into the logit the fixed values of variables other than x_k

Excerpt from Hamilton Table 7.4

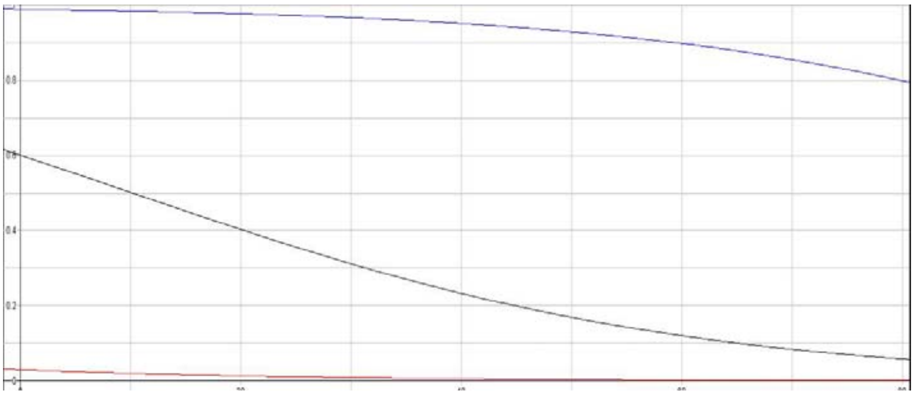
	B	S.E.	Wald	df	Sig.	Exp(B)	Minimum	Maximum	Mean
lived	-,040	,015	6,559	1	,010	,961	1,00	81,00	19,2680
educ	-,197	,093	4,509	1	,034	,821	6,00	20,00	12,9542
contam	1,299	,477	7,423	1	,006	3,664	,00	1,00	,2810
hsc	2,279	,490	21,591	1	,000	9,763	,00	1,00	,3072
nodad	-1,731	,725	5,696	1	,017	,177	,00	1,00	,1699
Constant	2,182	1,330	2,692	1	,101	8,866			

Logit:

L = 2.182 -0.04*lived -0.197*educ +1.299*contam +2.279*hsc -1.731*nodad

Here we let ”lived” vary and set in reasonable values for other variables

Conditional effect plot from Hamilton table 7.4 (fig7.5): effect of living for a long time in town

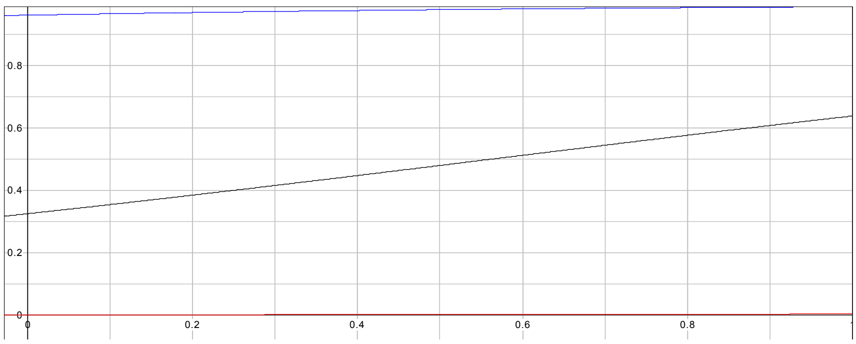


$y=1/(1+\exp(-(2.182-0.04x-0.197\times 12.95+1.299\times 0.28+2.279\times 0.31-1.731\times 0.17)))$ Mean

$y=1/(1+\exp(-(2.182-0.04x-0.197\times 6+1.299\times 1+2.279\times 1-1.731\times 0)))$ Max

$y=1/(1+\exp(-(2.182-0.04x-0.197\times 20+1.299\times 0+2.279\times 0-1.731\times 1)))$ Min

Conditional effect plot from Hamilton table 7.4 (fig7.6): effect of pollution on own land



$y=1/(1+\exp(-(2.182-0.04\times 19.27-0.197\times 12.95+1.299x+2.279\times 0.31-1.731\times 0.17)))$ Mean

$y=1/(1+\exp(-(2.182-0.04\times 1-0.197\times 6+1.299x+2.279\times 1-1.731\times 0)))$ Max

$y=1/(1+\exp(-(2.182-0.04\times 81-0.197\times 20+1.299x+2.279\times 0-1.731\times 1)))$ Min

Coefficients of determination

- Logistic regression does not provide measures comparable to the coefficient of determination in OLS regression
- Several measures analogous to R^2 have been proposed
- They are often called pseudo R^2
- Hamilton uses Aldrich and Nelson's

  pseudo $R^2 = \chi^2/(\chi^2+n)$

  where χ^2 = test statistic for the test of the whole model against a model with just a constant and n= the number of cases

Some pseudo R^2 in SPSS

- SPSS reports Cox and Snell, Nagelkerke, and in multinomial logistic regression also McFadden's proposal for R^2
- Aldrich and Nelson's pseudo R^2 can easily be computed by ourselves [pseudo $R^2 = \chi^2/(\chi^2+n)$]

Model Summary

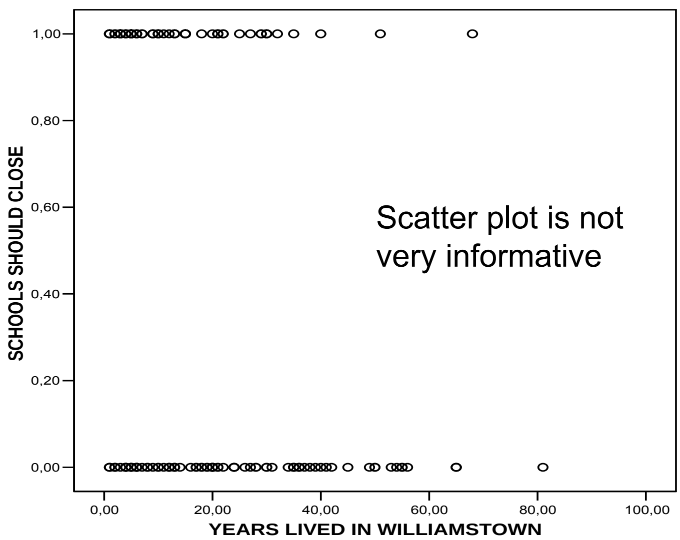
Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	***	***	***

Pseudo R-Square	
Cox and Snell	***
Nagelkerke	***
McFadden	***

Statistical problem: linearity of the logit

- Curvilinearity of the logit can give biased parameter estimates
- Scatter plot for y - x is not informative since y only has 2 values
- To test if the logit is linear in an x-variable one may do as follows
  - Group the x variable
  - For every group find average of y and compute the logit for this value
  - Make a graph of the logits against the grouped x

Y=”Closing school” vs. x= ”Years lived in town”



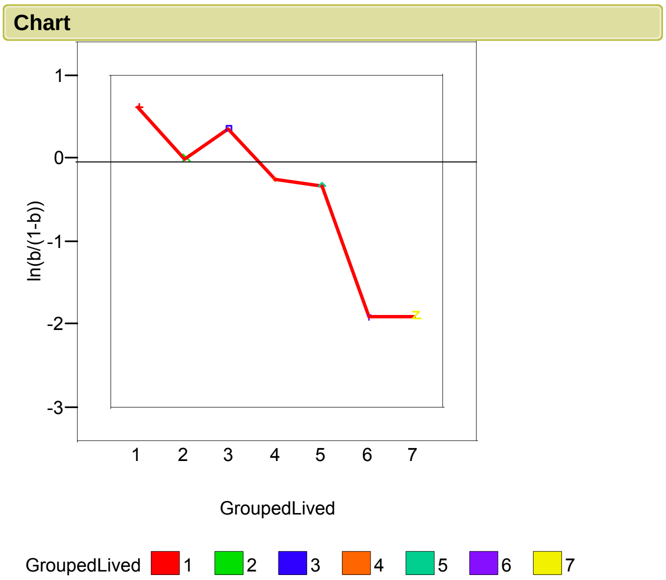
Linearity in logit: example

Recall: Logit = $L_i = \ln(O_i) = \ln\{p_i/(1-p_i)\}$

SCHOOLS SHOULD CLOSE		YEARS LIVED IN WILLIAMSTOWN (Banded)						
		<= 3	4-6	7-11	12-22	23-33	34-44	45+
N	OPEN	7	14	7	22	11	13	13
N	CLOSE	13	14	10	17	8	2	2
Within group	Mean (=p)	,65	,50	,59	,44	,42	,13	,13
Logit	Ln(p/(1-p))	0,619	0	0,364	-0,241	-0,323	-1,901	-1,901

Is the logit linear in ”years lived in town”?

Maybe!



In case of curvilinearity the odds ratio is non-constant

Assume the logit is curvilinear in education. Then the odds ratio for answering yes, adding one year of education, is:

$$\frac{e^{b_0+b_a*Alder+b_k*Kvinne+b_{utd}*(E.utd+1)+b_{utd2}*(E.utd+1)^2}}{e^{b_0+b_a*Alder+b_k*Kvinne+b_{utd}*E.utd+b_{utd2}*E.utd^2}}=$$

$$\frac{e^{b_{utd}+b_{utd2}*(E.utd^2+2E.utd+1)}}{e^{b_{utd2}*E.utd^2}}=\frac{e^{b_{utd}+b_{utd2}*(2E.utd+1)}}{e^0}=e^{b_{utd}+b_{utd2}*(2E.utd+1)}$$

Statistical problems: influence

- Influence from outliers and unusual x-values are just as problematic in logistic regression as in OLS regression
- Transformation of x-variables to symmetry will minimize the influence of extreme variable values
- Large residuals are indicators of large influence

Influence: residuals

- There are several ways to standardize residuals
  - ”Pearson residuals”
  - ”Deviance residuals”
- Influence can be based on
  - Pearson residual
  - Deviance residual
  - Leverage (potential for influence): i.e. the statistic h_j

Diagnostic graphs

Outlier plots can be based on plots of estimated probability of Y_i=1 (estimated P_i) against

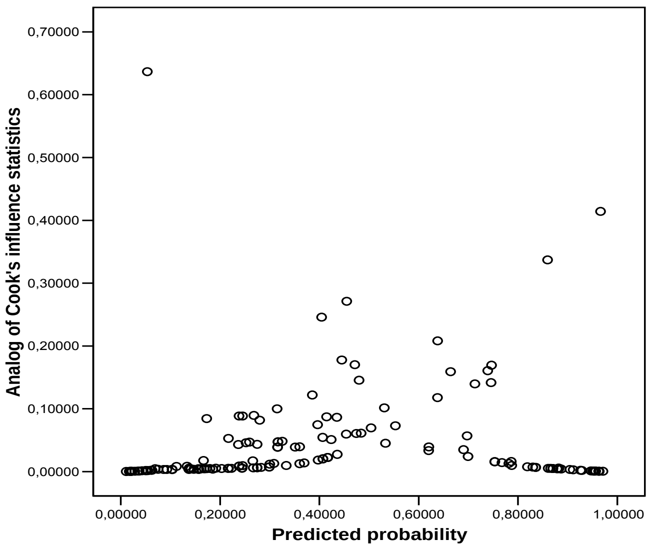
- Delta* B , ΔB_j , or
- Delta* Pearson Chisquare, $\Delta \chi^2_{P(j)}$, or
- Delta* Deviance Chisquare, $\Delta \chi^2_{D(j)}$

* “Delta” can be translated as “change in”

SPSS output

- **Cook's = delta B in Hamilton**
  - The logistic regression analogue of Cook's influence statistic. A measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients.
- **Leverage Value = h in Hamilton**
  - The relative influence of each observation on the model's fit.
- **DfBeta(s)** is not used by Hamilton in logistic regression
  - The difference in beta value is the change in the regression coefficient that results from the exclusion of a particular case. A value is computed for each term in the model, including the constant.

Delta B



SPSS output from ”Save” (1)

- **Unstandardized Residuals**
  - The difference between an observed value and the value predicted by the model.
- **Logit Residual**

$$\tilde{e}_i = \frac{e_i}{\hat{\pi}_i(1-\hat{\pi}_i)}; where\ e_i = y_i - \hat{\pi}_i$$

π_i is the probability that $y_i = 1$; the “hat” means estimated value

SPSS output from ”Save” (2)

- **Standardized = Pearson residual**
  - The command ”standardized” will make SPSS write a variable called ZRE_1 and labelled “Normalized residual”
  - This is the same as the Pearson residual in Hamilton
- **Studentized = [SQRT(delta deviance chisquare)]**
  - The command ”Studentized” will make SPSS write a variable called **SRE_1** and labelled “Standardized residual”
  - This is the same as the square root of ”delta Deviance chisquare” in Hamilton, i.e. ”delta Deviance chisquare” = $(SRE\_1)^2$
- **Deviance = Deviance residual**
  - The command ”Deviance” will make SPSS write a variable called **DEV_1** and labelled “Deviance value”
  - This is the same as the deviance residual in Hamilton

Computation of $\Delta\chi^2_{P(i)}$

- Based on the quantities provided by SPSS we can compute ”delta Pearson chisquare”
- Where it says r_j in the formula we put in ZRE_1 and where it says h_j we put in LEV_1

$$\Delta\chi^2_{P(j)} = \frac{r_j^2}{\left(1-h_j\right)}$$

Computation of $\Delta\chi^2_{D(i)}$

Based on the quantities provided by SPSS we can compute ”Delta Deviance Chisquare”

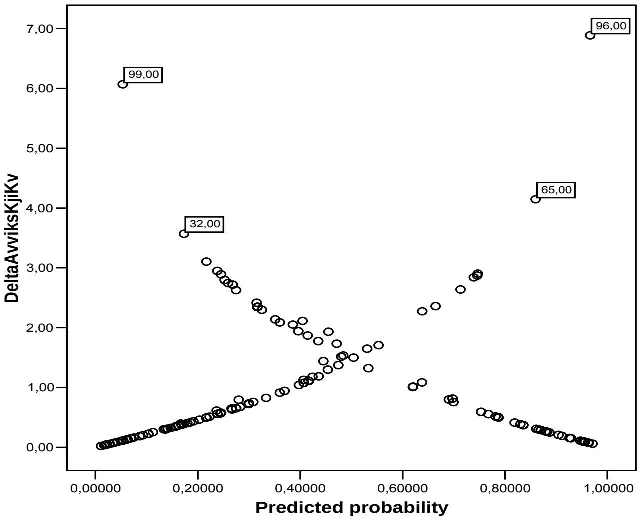
1. To find ”delta deviance chisquare” we square SRE_1

$$\Delta\chi^2_{D(j)} = SRE\_1 * SRE\_1$$

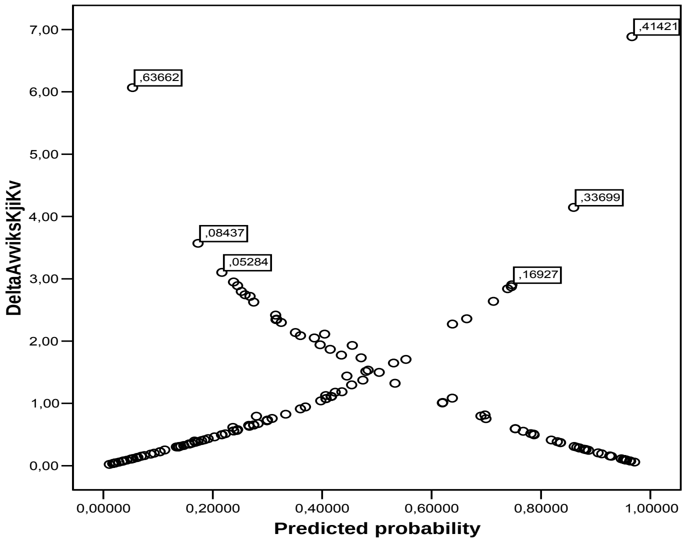
2. Alternatively we put in d_j=DEV_1 and h_j=LEV_1 in the formula

$$\Delta\chi^2_{D(j)} = \frac{d_j^2}{\left(1-h_j\right)}$$

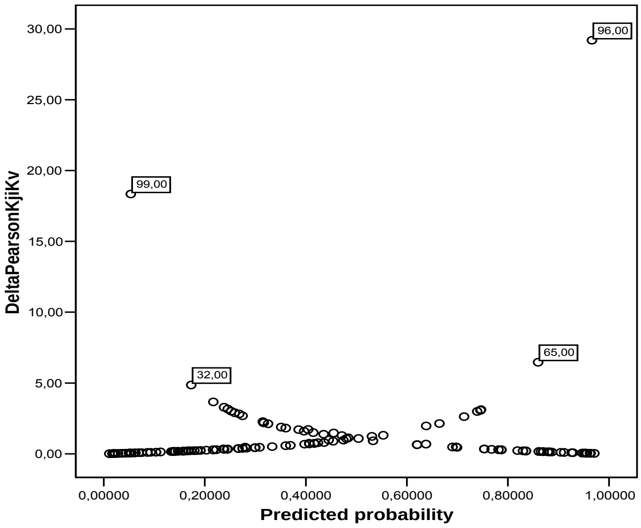
DeltaDevianceChisquare (with/CaseNO)



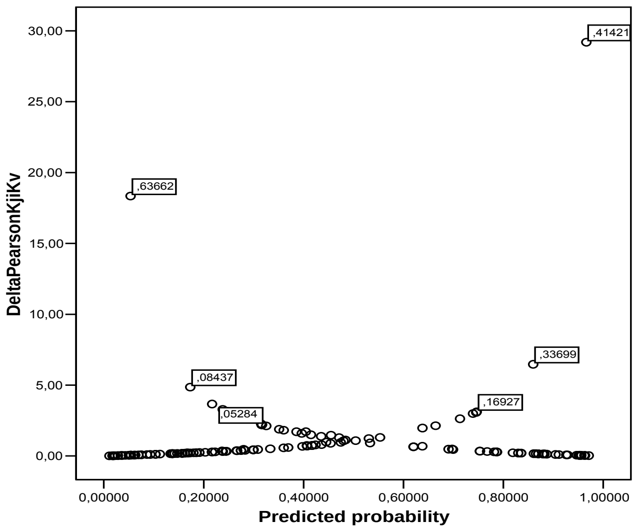
DeltaDevianceChisquare (with/delta B)



Delta Pearson Chisquare (with/CaseNO)



Delta Pearson Chisquare (with/ delta B)



Cases with large influence

Variables	Case No 96	Case No 65	Case No 99
Y=close	**1,00**	**,00**	**,00**
lived	68,00	40,00	1,00
educ	12,00	12,00	12,00
contam	,00	1,00	1,00
hsc	,00	1,00	1,00
nodad	,00	,00	,00
PRE_1	**,05**	**,86**	**,97**
COO_1	,64	,34	,41
RES_1	,95	-,86	-,97
SRE_1	2,46	-2,04	-2,62

Variables	Case No 96	Case No 65	Case No 99
ZRE_1	4,21	-2,48	-5,36
DEV_1	2,42	-1,98	-2,61
DFB0_1	-,32	,01	-,36
DFB1_1	,01	,00	,00
DFB2_1	,02	,01	,02
DFB3_1	-,08	-,15	-,18
DFB4_1	-,06	-,17	-,19
DFB5_1	-,08	,16	,14
DeltaPearsonKjiKv	18,34	6,47	29,20
DeltaAvviksKjiKv	6,07	4,14	6,89

From Cases to Patterns

- The figures shown previously are not identical to those you see in Hamilton
- Hamilton has corrected for the effect of identical patterns

Influence from a shared pattern of x-variables

- In a logistic regression with few variables many cases will have the same value on all x-variables. Every combination of x-variable values is called a pattern
- When many cases have the same pattern, every case may have a small influence, but collectively they may have unusually large influence on parameter estimates
- Influential patterns in x-values can give biased parameter estimates

Influence: Patterns in x-values

- Predicted value, and hence the residual will be the same for all cases with the same pattern
- Influence from pattern j can be found by means of
  - The frequency of the pattern
  - Pearson residual
  - Deviance residual
  - Leverage: i.e. the statistic h_j

Finding X-pattern by means of SPSS

- In the ”Data” – menu find the command ”Identify duplicate cases”
- Mark the x-variables that are used in the model and move them to ”Define matching cases by”
- Cross for ”Sequential count of matching cases in each group” and ”Display frequencies for created variables”
- This produces two new variables. One, ”MatchSequence”, numbers cases sequentially 1, 2, … where several patterns are identical. If the pattern is unique this variable has the value 0.
- The other variable, ”Primary…”, has the value 0 for duplicates and 1 for unique patterns

X-patterns in SPSS; Hamilton p238-242

	Frequency	Percent	Valid Percent	Cumulative Percent
Duplicate Case	21	13,7	13,7	13,7
Primary Case	132	86,3	86,3	100,0
Total	**153**	**100,0**	**100,0**	

Sequential count of matching cases	Frequency	Percent	Valid Percent	Cumulative Percent
0 [115 patterns with 1 case]	115	75,2	75,2	75,2
1 [17 patterns with 2 or 3 cases]	17	11,1	11,1	86,3
2 [17–4=13 patterns with 2 cases]	17	11,1	11,1	97,4
3 [4 patterns with 3 cases]	4	2,6	2,6	100,0
Total	**153**	**100,0**	**100,0**	

Hamilton table 7.6 Symbols

J	# unique patterns of x-values in the data (J<=n)
m_j	# cases with the pattern j (m>=1)
$\hat{P}_j$	Predicted probability of Y=1 for case with pattern j
Y_j	Sum of y-values for cases with pattern j (= # cases with pattern j and y=1)
r_j	Pearson residual for pattern j
χ^2_P	Pearson Chisquare statistic
d_j	Deviance residual for pattern j
χ^2_D	Deviance Chisquare statistic
h_i	Leverage for case i
h_j	Leverage for pattern j

New values for $\Delta\chi^2_{P(i)}$ and $\Delta\chi^2_{D(i)}$

- By ”Compute” one may calculate the Pearson residual (equation 7.19 in Hamilton) and delta Pearson chisquare (equation 7.24 in Hamilton) once more. This will provide the correct values
- The same applies for deviance residual (equation 7.21) and delta deviance chisquare (equation 7.25a)

Leverage and residuals (1)

- Leverage of a pattern is obtained as number of cases with the pattern times the leverage of a case with this pattern. The leverage of a case is the same as in OLS regression
- $h_j = m_j*h_i$
- Pearson residual can be found from

$$r_j = \frac{Y_j - m_j \hat{P}_j}{\sqrt{m_j \hat{P}_j \left(1 - \hat{P}_j\right)}}$$

Leverage and residuals (2)

- Deviance residual can be found from

$$d_j = \pm\sqrt{\left\{2\left[Y_j \ln\left(\frac{Y_j}{m_j \hat{P}_j}\right) + \left(m_j - Y_j\right)\ln\left(\frac{m_j - Y_j}{m_j\left(1 - \hat{P}_j\right)}\right)\right]\right\}}$$

Two Chi-square statistics

- Pearson Chi-square statistics

$$\chi_P^2 = \sum_{j=1}^{J} r_j^2$$

- Deviance Chi-square statistics

$$\chi_D^2 = \sum_{j=1}^{J} d_j^2$$

- Equations are the same for both cases and patterns

The Chisquare statistics

Both Chisquare statistics:

1. Pearson-Chisquare χ^2_P and
2. Deviance-Chisquare χ^2_D

- Can be read as a test of the null hypothesis of no difference between the estimated model and a “saturated model”, that is a model with as many parameters as there are cases/ patterns

Large values of measures of influence

- Measures of influence based on changes in (Δ) the statistic/ parameter value due to excluded cases with pattern j
  - ΔB_j "delta B" - analogue to Cook's D
  - $\Delta\chi^2_{P(i)}$ "delta Pearson-Chisquare"
  - $\Delta\chi^2_{D(i)}$ "delta Deviance-Chisquare"

What is a large value of $\Delta\chi^2_{P(i)}$ and $\Delta\chi^2_{D(i)}$

- Both $\Delta\chi^2_{P(i)}$ and $\Delta\chi^2_{D(i)}$ measure how badly the model fits the pattern j. Large values indicates that the model would fit the data much better if all cases with this pattern were excluded
- Since both measures are distributed asymptotically as the chisquare distribution, values larger than 4 indicate that a pattern affects the estimated parameters "significantly"

ΔB_j "delta B"

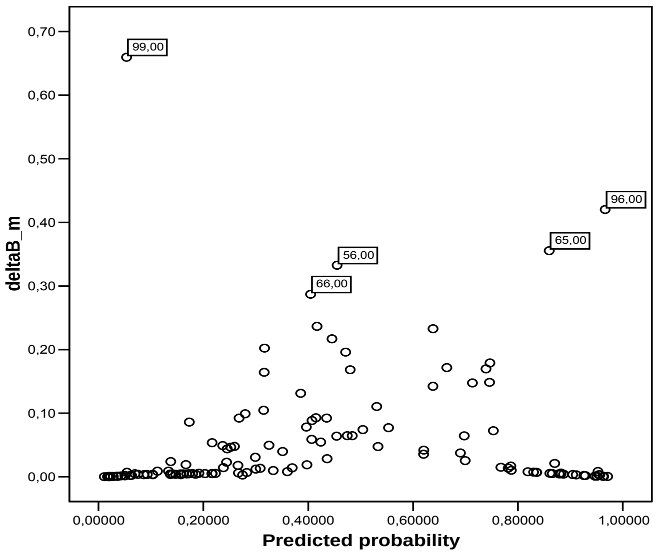
- Measures the standardized change in the estimated parameters (b_k) that obtain when all cases with a given pattern j are excluded

$$\Delta B_j = \frac{r_j^2 h_j}{\left(1 - h_j\right)^2}$$

Larger values means larger influence

ΔB_j >= 1 must in any case be seen as "large influence"

delta B (with caseNO)

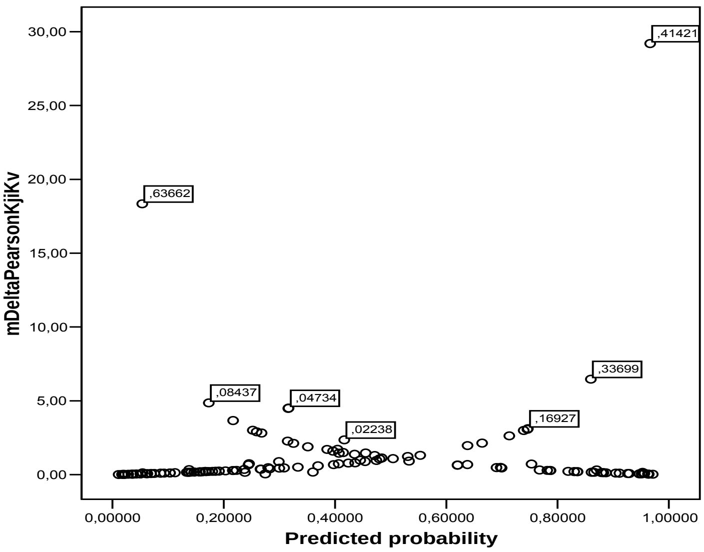


$\Delta\chi^2_{P(i)}$ "Delta Pearson Chisquare"

- Measures the reduction in Pearson χ^2 that obtains from excluding all cases with pattern j

$$\Delta\chi^2_{P(j)} = \frac{r_j^2}{\left(1-h_j\right)}$$

Delta Pearson Chisquare (with delta B)



$\Delta\chi^2_{D(i)}$ “Delta Deviance Chisquare”

- Measures changes in deviance that obtains from excluding all cases with pattern j

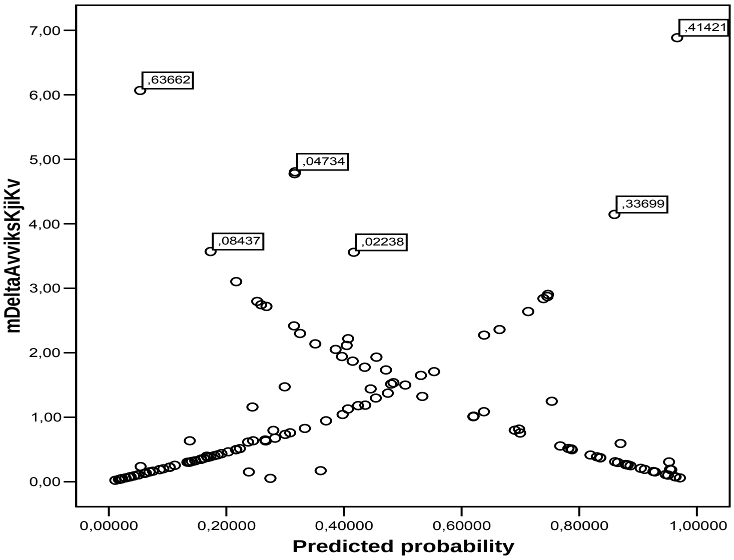
$$\Delta\chi^2_{D(j)} = \frac{d_j^2}{\left(1-h_j\right)}$$

- This is equivalent to

$$\Delta\chi^2_{\mathcal{D}(j)} = -2\left[\mathcal{LL}_{\mathrm{K}} - \mathcal{LL}_{\mathrm{K(j)}}\right]$$

$\mathcal{LL}_{\mathbf{K}}$ is the LogLikelihood of a model with K parameters estimated on the whole sample and $\mathcal{LL}_{\mathbf{K(j)}}$ is from the estimate of the same model when all cases with pattern j are excluded

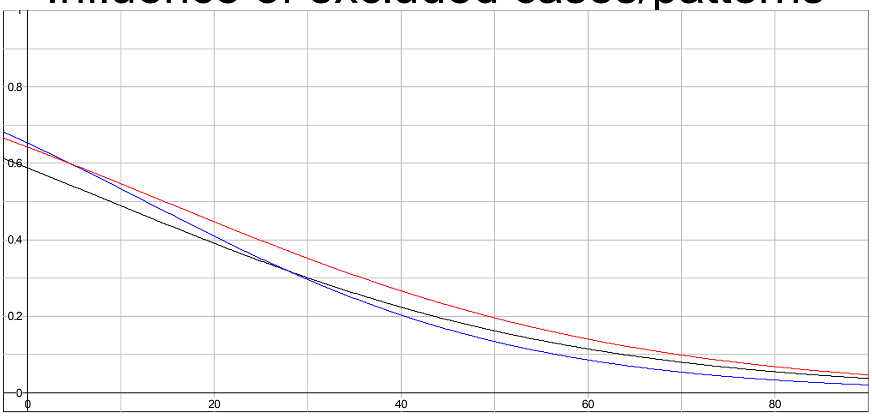
Delta Deviance Chisquare (with delta B)



Influence of excluded cases/patterns

	Logit coefficient		
Variables in the model	Sample	Excluding case 99 $\Delta\chi^2 P(i)$ =18,34	Excluding case 96 $\Delta\chi^2 P(i)$ =29,20
lived	-,040	-,045	-,052
educ	-,197	-,224	-,214
contam	1,299	1,490	1,382
hsc	2,279	2,492	2,347
nodad	-1,731	-1,889	-1,658
Constant	2,182	2,575	2,530
2*LL(modell)	**-142,652**	**-135,425**	**-136,124**

Influence of excluded cases/patterns



y=1/(1+exp(-(2.18-0.04x-0.2×13+1.3×0.28+2.28×0.31-1.73×0.17)))
y=1/(1+exp(-(2.53-0.05x-0.21×13+1.38×0.28+2.35×0.31-1.65×0.17)))
y=1/(1+exp(-(2.58-0.04x-0.22×13+1.49×0.28+2.49×0.31-1.89×0.17)))

Conclusions (1)

Ordinary OLS do not work well for
dichotomous dependent variables since

- It is impossible to obtain normally distributed errors or homoscedasticity, and since
- The model predicts probabilities outside the interval [0-1]

The Logit model is for theoretical reasons better

- Likelihood ratio tests statistic can be used to test nested models analogous to the F-statistic
- In large samples the chisquare distributed Wald statistic [or the normally distributed t=SQRT(Wald)] will be able to test single coefficients and provide confidence intervals
- There is no statistic similar to the coefficient of determination

Conclusions (2)

- Coefficient of estimated models can be interpreted by
  1. Log-odds (direct interpretation)
  2. Odds
  3. Odds ratio
  4. Probability (conditional effect plot)
- Non-linearity, case with influence, and multicollinearity leads to the same kinds of problems as in OLS regression (inaccurate or uncertain parameter values)
- Discrimination leads to problems of uncertain parameter values (large variance estimates)
- Diagnostic work is important