

Conditional Logistic Regression

Azmi Mohd Tamil

Extension of Binary Logistic Regression

- Conditional logistic regression is the extension for the (binary) logistic regression when the categorical dependent outcome has only two levels (case & control) but are matched e.g. by gender & age. For example, in a matched case-control study matched by gender & age.
- If it is unmatched, we just use chi-square and multiple logistic regression.
- If it is matched, we use McNemar and conditional logistic regression.

Conditional LogitR in SPSS?

- Not possible to do Conditional Logistic Regression directly in SPSS. Instead you have to format the data accordingly and use either;
 - Multinomial logistic regression (matched 1 to 1) or
 - Cox Regression (matched 1 to 1 or 1 to many)
- So we will show you how to format the data in both situations.

Conditional logistic regression

Using Multinomial Logistic Regression

Formating for Multinomial Command

Table Va. First five cases of a Matched case-control study: outcomes.

Dead					Alive				
x1	x2	x3	diabetes	race	x1	x2	x3	diabetes	race
84.00	82.10	45.00	1	1	84.0	73.2	47.0	0	1
83.10	86.40	52.00	1	3	86.1	81.1	51.0	1	1
84.60	87.00	53.00	1	2	87.3	76.8	48.0	0	3
84.00	78.80	50.50	0	2	84.2	71.4	48.5	1	1
83.50	88.20	46.00	1	2	83.2	73.7	47.0	1	1



Table Vb. First five cases of a Matched case-control study: variables.

Outcome	x1_diff	x2_diff	x3_diff	Diabetes diff	Malay dead	Indian dead	Malay alive	Indian alive	Malay diff	Indian diff
1	0.00	8.90	-2.00	1.00	0.00	0.00	0.00	0.00	0.00	0
1	-3.00	5.30	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1
1	-2.70	10.20	5.00	1.00	1.00	0.00	0.00	1.00	1.00	-1
1	-0.20	7.40	2.00	-1.00	1.00	0.00	0.00	0.00	1.00	0
1	0.30	14.50	-1.00	0.00	1.00	0.00	0.00	0.00	1.00	0

For Diff; It is “Case – Control = ?”

Table Vb. First five cases of a Matched case-control study: variables.

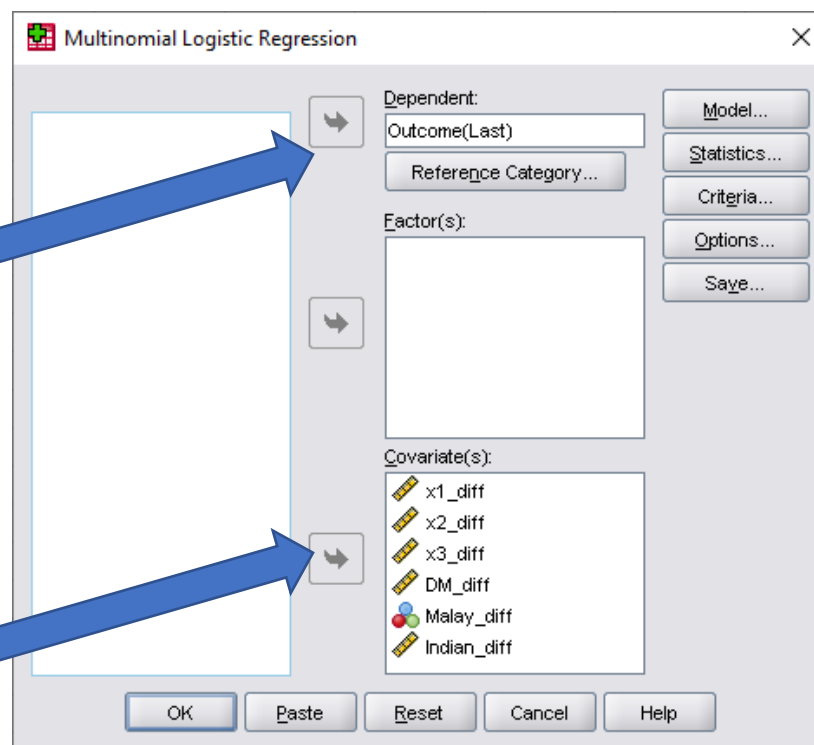
Outcome	x1_diff	x2_diff	x3_diff	Diabetes diff	Malay dead	Indian dead	Malay alive	Indian alive	Malay diff	Indian diff
I	0.00	8.90	-2.00	1.00	0.00	0.00	0.00	0.00	0.00	0
I	-3.00	5.30	1.00	0.00	0.00	1.00	0.00	0.00	0.00	1
I	-2.70	10.20	5.00	1.00	1.00	0.00	0.00	1.00	1.00	-1
I	-0.20	7.40	2.00	-1.00	1.00	0.00	0.00	0.00	1.00	0
I	0.30	14.50	-1.00	0.00	1.00	0.00	0.00	0.00	1.00	0

<http://www.palmx.org/drtamil/spss/conditional/yhchan-condreg-multinomial.sav>

	Outcome	x1_diff	x2_diff	x3_diff	DM_diff	Malay_diff	Indian_diff
1	1	0.0	8.9	-2	1	0	0
2	1	-3.0	5.3	1	0	0	1
3	1	-2.7	10.2	5	1	1	-1
4	1	-0.2	7.4	2	-1	1	0
5	1	0.3	14.5	-1	0	1	0

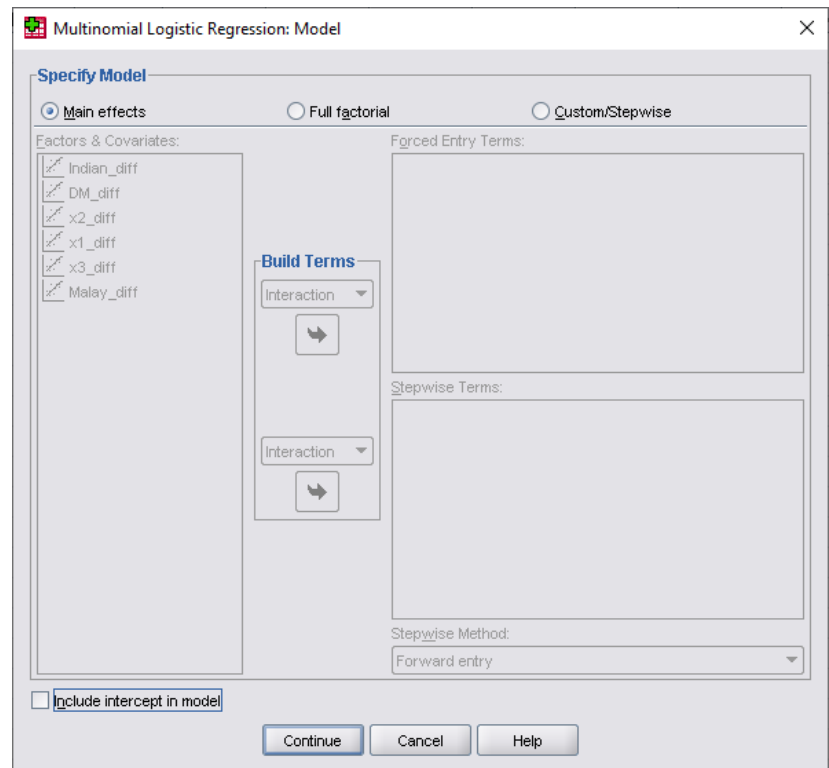
Analysis CLogitR Using Multinomial

- To perform the analysis in SPSS, go to Analyse, Regression & select Multinomial logistic.
- Put the Outcome variable into the Dependent option.
- And all the difference variables computed earlier into the Covariates option.



Analysis CLogitR Using Multinomial 2

- Click on the Model folder, have to uncheck the “Include Intercept in model” option. Let us use the main effects model.



Analysis CLogitR Using Multinomial 3

1. These are not valid for matched case-control studies (not required).

- goodness-of-fit statistics and
- the classification table

2. These are valid and interpreted as usual.

- Model fitting information,
- the likelihood ratio and
- R-square statistics

Multinomial Logistic Regression: Statistics

☒ Case processing summary

Model

☒ Pseudo R-square ☐ Cell probabilities

☐ Step summary ☐ Classification table

☒ Model fitting information ☐ Goodness-of-fit

☐ Information Criteria ☐ Monotonicity measures

Parameters

☒ Estimates Confidence Interval (%): 95

☒ Likelihood ratio tests

☐ Asymptotic correlations

☐ Asymptotic covariances

Define Subpopulations

☒ Covariate patterns defined by factors and covariates

☐ Covariate patterns defined by variable list below

Subpopulations:

Indian_diff DM_diff x2_diff Malay_diff x3_diff x1_diff x2_diff DM_diff Indian_diff

Continue Cancel Help

Analysis CLogitR Using Multinomial 4

- Table Vc shows the message that a conditional logistic regression is being performed after you run the analysis.

Table Vc. Conditional logistic regression message.

Warning
The dependent variable has only one valid value. A conditional logistic regression model will be fitted.

Analysis CLogitR Using Multinomial 5

- Table Vd shows the parameter estimates.

Table Vd. Parameter estimates: conditional logistic.

Parameter estimates									
Outcome		B	Std. error	Wald	df	Sig.	Exp(B)	95% confidence interval for Exp(B)	
								Lower bound	Upper bound
Dead	x1_diff	.021	.078	.071	1	.789	1.021	.877	1.188
	x2_diff	.400	.086	21.496	1	.000	1.492	1.260	1.767
	x3_diff	-.499	.143	12.088	1	.001	.607	.458	0.804
	diabetes_diff	1.065	.521	4.188	1	.041	2.902	1.046	8.049
	Malay_diff	-.178	.629	.080	1	.777	.837	.244	2.871
	Indian_diff	-.112	.647	.030	1	.862	.894	.252	3.173

Practical CondLogitR Using Multinomial

- Download data from <http://www.palmx.org/drtamil/spss/conditional/Malaria-CondLogRegres-Multinomial.sav>
- Do McNemar for all the paired data between Case & Control.
- Identify all the significant variables and include them in the CondLogitR analysis.

Interpret the Significant Results

		Parameter Estimates							
Malaria		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
1	Non-Malaysian	.927	.778	1.419	1	.234	2.527	.550	11.622
	Low Education Level	.272	.790	.118	1	.731	1.313	.279	6.175
	High Risk Occupation (Army, Lumberjacks, Plantation workers)	1.377	.684	4.049	1	.044	3.964	1.036	15.162
	Low Income Group	.284	.766	.137	1	.711	1.328	.296	5.960
	No TV Ownership	-1.143	.934	1.497	1	.221	.319	.051	1.990
	No Motorcycle Ownership	-.131	.659	.040	1	.842	.877	.241	3.193
	No Car Ownership	-.765	.662	1.335	1	.248	.465	.127	1.704
	Low Self-Perception of Risk	1.854	.789	5.515	1	.019	6.383	1.359	29.984
	Don't Use Treated Bed Net	.188	.631	.087	1	.768	1.204	.350	4.148
	No MOH Residual Spraying of Home	1.302	.641	4.130	1	.042	3.676	1.047	12.898
	Owner Did Not Spray Aerosol at Home	.923	.771	1.431	1	.232	2.517	.555	11.414
	Travel Last 2 Weeks into Jungle/Plantation	1.780	.851	4.377	1	.038	5.929	1.119	31.410
	Home outside village	.827	.911	.825	1	.364	2.287	.384	13.638
	Home near jungle	.720	.697	1.067	1	.302	2.054	.524	8.052
	Home near water bodies	.907	.934	.943	1	.331	2.477	.397	15.453
	Opening near roof eaves	-.161	.811	.040	1	.842	.851	.174	4.173
	Roof not made of tiles	.205	.952	.048	1	.830	1.227	.190	7.928
	Home no ceiling	1.155	.734	2.475	1	.116	3.175	.753	13.394

Conditional logistic regression

Using Cox Regression

Formating for Cox Regression Command

Table Va. First five cases of a Matched case-control study: outcomes.

Dead					Alive				
x1	x2	x3	diabetes	race	x1	x2	x3	diabetes	race
84.00	82.10	45.00	1	1	84.0	73.2	47.0	0	1
83.10	86.40	52.00	1	3	86.1	81.1	51.0	1	1
84.60	87.00	53.00	1	2	87.3	76.8	48.0	0	3
84.00	78.80	50.50	0	2	84.2	71.4	48.5	1	1
83.50	88.20	46.00	1	2	83.2	73.7	47.0	1	1



Table VIa. Conditional logistic regression (1:1 matching) using Cox regression option.

Matching number	x1	x2	x3	Dead	Diabetes	Race	Time
1	84.0	82.1	45.0	1	1	1	1
1	84.0	73.2	47.0	0	0	1	2
2	83.1	86.4	52.0	1	1	3	1
2	86.1	81.1	51.0	0	1	1	2
3	84.6	87.0	53.0	1	1	2	1
3	87.3	76.8	48.0	0	0	3	2

Status = 1 & 0; Time = 1 & 2

Table VIa. Conditional logistic regression (1:1 matching) using Cox regression option.

Matching number	x1	x2	x3	Dead	Diabetes	Race	Time
1	84.0	82.1	45.0	1	1	1	1
1	84.0	73.2	47.0	0	0	1	2
2	83.1	86.4	52.0	1	1	3	1
2	86.1	81.1	51.0	0	1	1	2
3	84.6	87.0	53.0	1	1	2	1
3	87.3	76.8	48.0	0	0	3	2

<http://www.palmx.org/drtamil/spss/conditional/yhchan-condreg-cox.sav>

	Pair	x1	x2	x3	Dead	DM	Race	Time
1	1	84.0	82.1	45	1	1	1	1
2	1	84.0	73.2	47	0	0	1	2
3	2	83.1	86.4	52	1	1	3	1
4	2	86.1	81.1	51	0	1	1	2
5	3	84.6	87.0	53	1	1	2	1
6	3	87.3	76.8	48	0	0	3	2

Status = 1 & 0; Time = 1 & 2

	Label	Status	Time
Case	Dead	1	1
Control	Alive	0	2

<http://www.palmx.org/drtamil/spss/conditional/yhchan-condreg-cox.sav>

	Pair	x1	x2	x3	Dead	DM	Race	Time
1	1	84.0	82.1	45	1	1	1	1
2	1	84.0	73.2	47	0	0	1	2
3	2	83.1	86.4	52	1	1	3	1
4	2	86.1	81.1	51	0	1	1	2
5	3	84.6	87.0	53	1	1	2	1
6	3	87.3	76.8	48	0	0	3	2

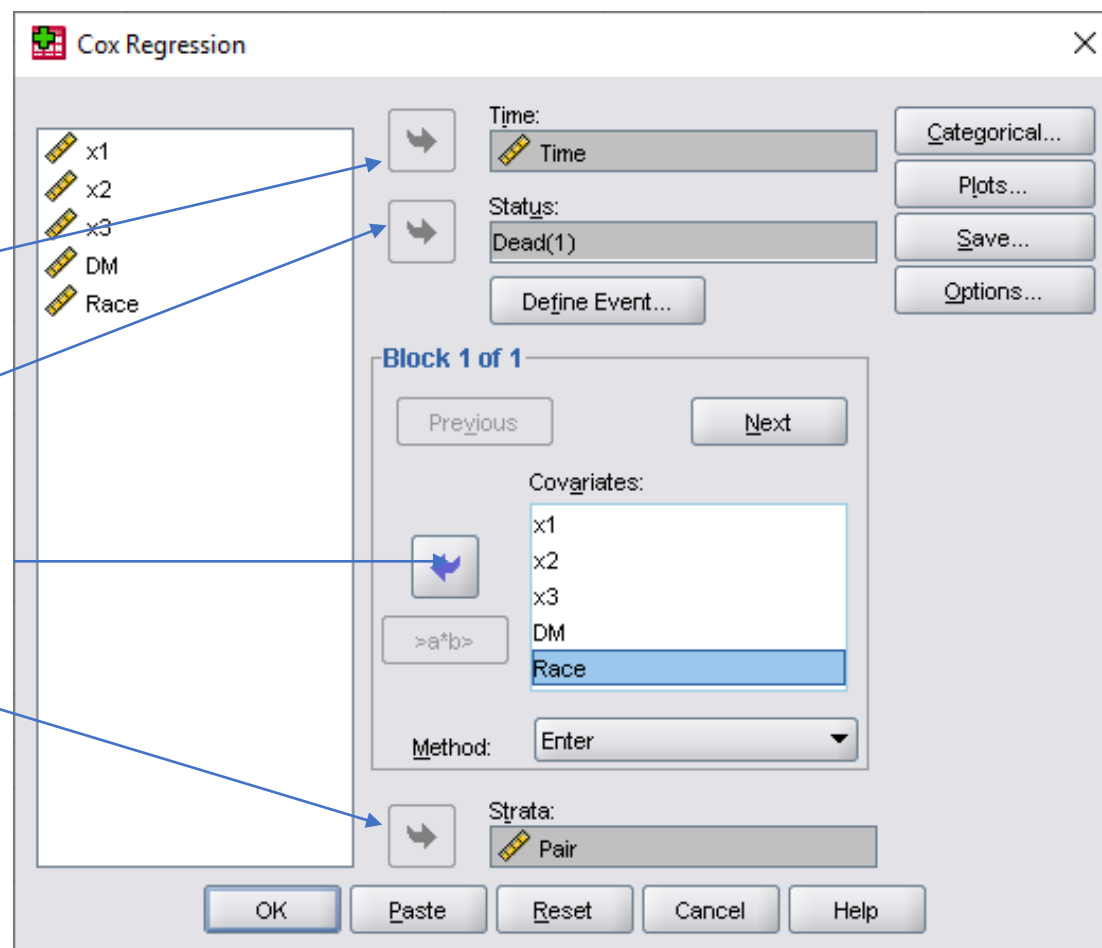
Status = 1 & 0; Time = 1 & 2

	Label	Status	Time
Case	Dead	1	1
Control	Alive	0	2

- We need a matching number to “link” the case and control. That matching number is the variable “Pair”.
- The Dead variable is the outcome status of the subject (dead = 1 and alive = 0).
- We need a variable Time as the response variable where the dead has a Time = 1 and the alive (censored) has Time = 2.

Analysis CLogitR Using Cox Regression

- To perform the analysis, in SPSS, go to Analyse, Survival, Cox Regression to get the following template.
- Put Time in the Time option.
- Put Dead in the Status option (define Event = 1).
- Put the variables of interest into the Covariates option.
- Lastly include the Match number (Pair) in the Strata option.



Analysis CLogitR Using Cox Regression

- This will produce exactly the same results in Table Vd.

Table Vd. Parameter estimates: conditional logistic.

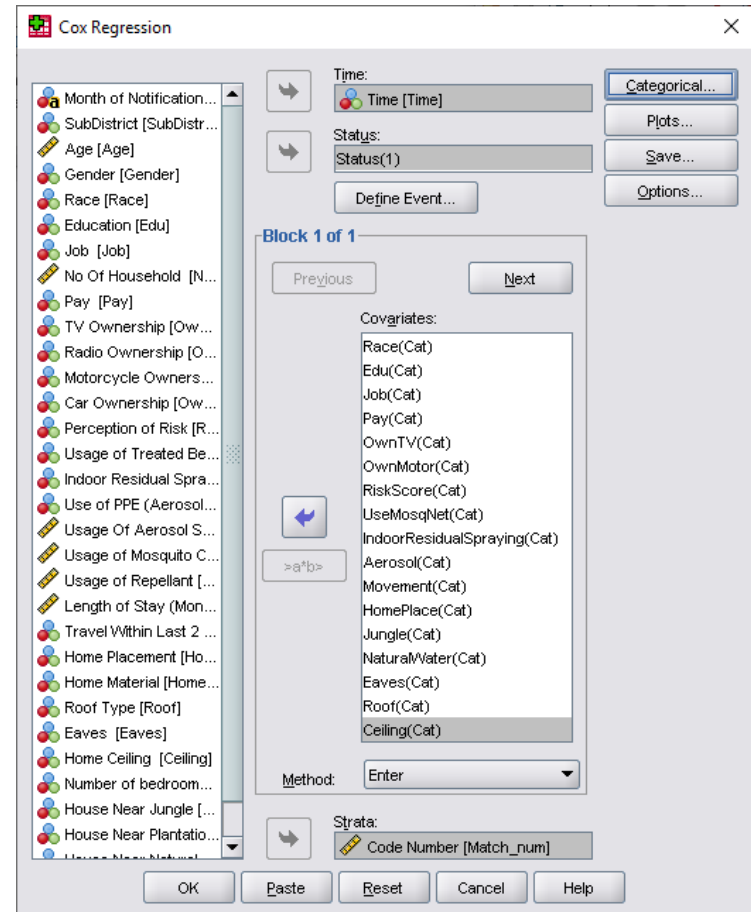
Parameter estimates									
Outcome		B	Std. error	Wald	df	Sig.	Exp(B)	95% confidence interval for Exp(B)	
								Lower bound	Upper bound
Dead	x1_diff	.021	.078	.071	1	.789	1.021	.877	1.188
	x2_diff	.400	.086	21.496	1	.000	1.492	1.260	1.767
	x3_diff	-.499	.143	12.088	1	.001	.607	.458	0.804
	diabetes_diff	1.065	.521	4.188	1	.041	2.902	1.046	8.049
	Malay_diff	-.178	.629	.080	1	.777	.837	.244	2.871
	Indian_diff	-.112	.647	.030	1	.862	.894	.252	3.173

Practical CondLogitR Using Cox Regression

- Download data from <http://www.palmx.org/drtamil/spss/conditional/Malaria-CondLogRegres-Cox.sav>
- We have done McNemar for all the paired data between Case & Control in earlier exercise.
- Include all the significant variables in the CondLogitR analysis.

Analysis CLogitR Using Cox Regression

- To perform the analysis, in SPSS, go to Analyse, Survival, Cox Regression to get the following template.
- Put Time in the Time option.
- Put Status in the Status option (define Event = 1).
- Put the variables of interest into the Covariates option. Click on Categorical and change Reference value to First for all.
- Lastly include the Match number (Pair) in the Strata option.



Interpret the Significant Results

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Race	.795	.765	1.080	1	.299	2.214	.494	9.920
Edu	.080	.780	.011	1	.918	1.083	.235	4.995
Job	1.209	.649	3.469	1	.063	3.351	.939	11.964
Pay	.116	.752	.024	1	.878	1.123	.257	4.899
OwnTV	-1.089	.925	1.387	1	.239	.336	.055	2.062
OwnMotor	.030	.654	.002	1	.963	1.031	.286	3.718
RiskScore	1.719	.787	4.772	1	.029	5.582	1.193	26.107
UseMosqNet	.179	.618	.084	1	.772	1.196	.356	4.013
IndoorResidualSpraying	1.369	.634	4.667	1	.031	3.931	1.135	13.610
Aerosol	.929	.751	1.529	1	.216	2.533	.581	11.044
Movement	1.865	.853	4.785	1	.029	6.458	1.214	34.347
HomePlace	.890	.889	1.002	1	.317	2.435	.426	13.905
Jungle	.709	.697	1.036	1	.309	2.032	.519	7.963
NaturalWater	.989	.918	1.160	1	.282	2.687	.445	16.247
Eaves	-.294	.802	.134	1	.714	.746	.155	3.589
Roof	.379	.950	.159	1	.690	1.461	.227	9.409
Ceiling	1.260	.720	3.064	1	.080	3.526	.860	14.456

Is the result the same as using Multinomial technique? If not the same, explain why.

Reference

- Chan YH, 2005. Biostatistics 305. Multinomial logistic regression. Singapore Med J. 2005 Jun;46(6):259-68; quiz 269. PMID: 15902353
- Thesis UKM – Marsudi Manah.