With Wisdom
We Explore

UTHM
Universiti Tun Hussein Onn Malaysia

Review on Poor Household Income in Malaysia using New Model Averaging

By

DR.KHUNESWARI A/P P.GOPAL PILLAY

Introduction

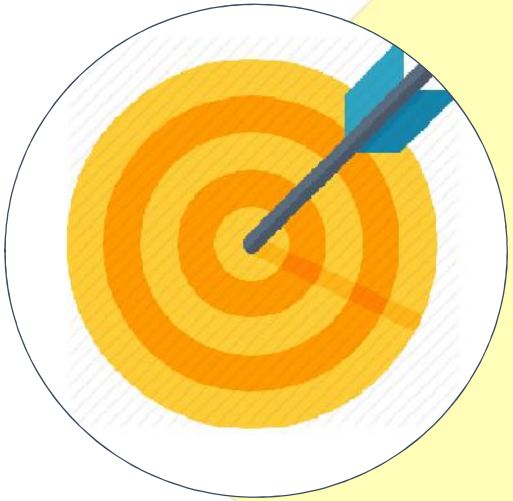
Model Building

- Model Averaging (MA) aims to overcome Model Selection (MS) issues regarding underestimation of parameter estimates.
- MA include all covariates being studied regardless of its significant in the final best model

Household Poverty

- Household poverty defined by Poverty Line Income (PLI) .
- Gross income under the national PLI will be categorized as poor.
- There are two aspects which are income/expenditure and secondly on non-income factors (ex-ample: household characteristic, health, state, amenities and etc.).

- ✓ New approach of MA (NMA) which is based on MA approach but with elimination of insignificant variables.
- ✓ Study the relationship between household characteristic and state with poverty to examine whether or not it effects the Malaysian's household poverty.



AIM

- **Ilustrate NMA procedures**
- **Highlight the cause of household poverty**
- **Combined data for year 2012 and 2016**

METHODOLOGY

Multiple Binary Logit

- Modelled data with **binomial outcome (values 0 and 1)**

$$P_i = \frac{exp^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_q X_q + \varepsilon}}{1 + exp^{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_q X_q + \varepsilon}}$$

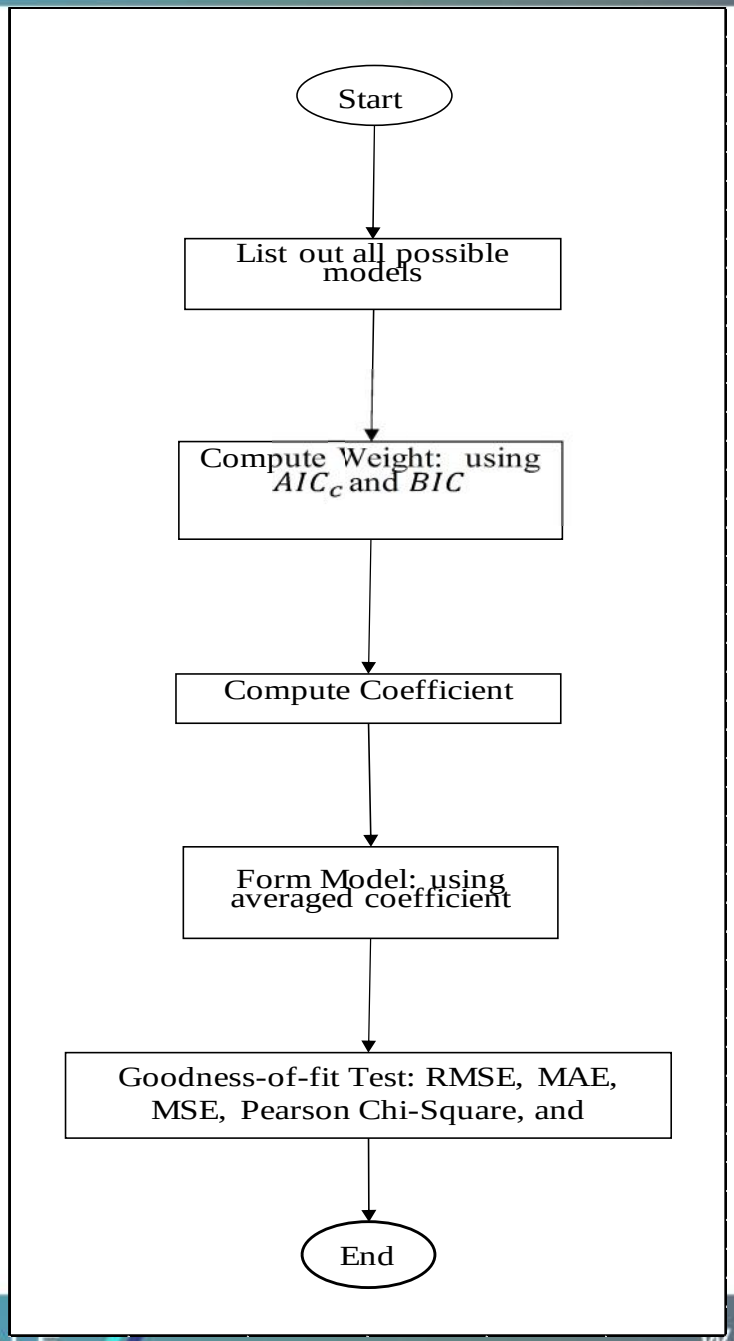
- **MBL** present the results of success/failure in forms of **probability**
- Example, a probability of 0.80 means that there is 80% chance of outcome 1 (success) to occur.

Variable Descriptions

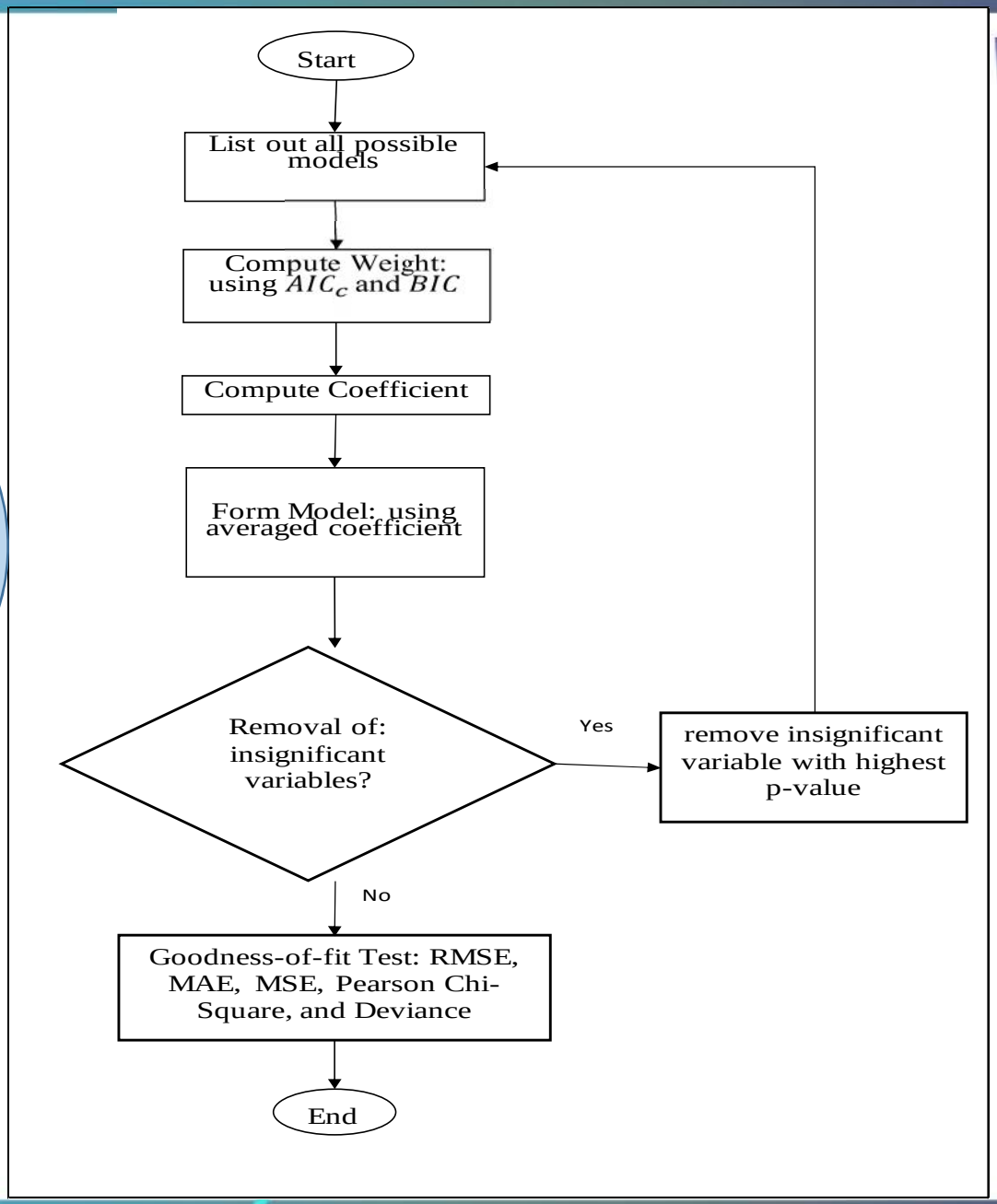
Variable	Description
Y	Poverty Level 0: Not Poor 1: Poor
X_1	State 01: Johor 02: Kedah 03: Kelantan 04: Melaka 05: Negeri Sembilan 06: Pahang 07: Pulau Pinang 08: Perak 09: Perlis 10: Selangor 11: Terengganu 12: Sabah 13: Sarawak 14: Kuala Lumpur 15: Labuan 16: Putrajaya
X_2	Household Age (H.Age) Age of the head of household
X_3	Household Gender (H. Gen) 1: Male 2: Female

Variable	Description
X_4	Household Marital (H. Mar) 1: Never married 2: Married 3: Widowed 4: Divorced 5: Separated
X_5	Household Activity (H. Act) 1: Employer 2: Government employee 3: Private employee 4: Own account worker 5: Unpaid family worker 6: Unemployed 7: Housewife 8: Student 9: Pensioner 10: Others 11: Child not at school
X_6	Household Size (H. Size) Total number of household member
X_7	Region 1: Peninsular Malaysia 2: Sabah (including Labuan) 3: Sarawak
X_8	Net Income (N. Inc) Net total is the total amount of income

Model Averaging Guidelines



New Model Averaging Guidelines



Step 1: All Possible Models

$$N = \sum_{j=1}^{q}((^{q}C_{j})) = \frac{q!}{j!\,(q-r)}$$

$$1(^{8}C_{1}) + 1(^{8}C_{2}) + (^{8}C_{3}) + 1(^{8}C_{4}) + 1(^{8}C_{5}) + 1(^{8}C_{6}) + 1(^{8}C_{7}) + 1(^{8}C_{8})$$

$$= 255\ possible\ models$$

Step 2: Weight Computation

$$W_m = \frac{\exp(\frac{I_m}{2})}{\sum_{m=1}^{M} \exp \frac{I_m}{2}}$$

I_m is the type of model selection criterion (AIC_c orBIC).

Possible Models	AIC_c	Weight
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_8 X_8$	-6.59	0.09
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_8 X_8$	-6.3	0.08
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8$	-6.27	0.08
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8$	-6.26	0.08
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_8 X_8$	-5.44	0.05
.	.	.
.	.	.
.	.	.
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4$	-0.33	0.00

Weights based on AICc

Possible Models	BIC	Weight
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_8 X_8$	16.45	0.52
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3$	18.13	0.23
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_4 X_4 + \hat{\beta}_8 X_8$	21.33	0.05
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_3 X_3 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8$	21.36	0.05
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_8 X_8$	22.19	0.03
.	.	.
.	.	.
.	.	.
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_7 X_7 + \hat{\beta}_8 X_8$	27.14	0.00

Weights based on BIC

Step 3: Coefficient Estimate

$$\widehat{\beta}_p = \sum_{m=1}^{M} w_m \widehat{\beta}_{(p,m)}$$

where $\widehat{\beta}_{(p,\,m)}$ is the estimate of β_p under model for $m = 1,2,\ldots, M$

$$\widehat{\beta}_0 = \frac{\beta_{(0,1)} И_1 + \beta_{(0,2)} W_2 + \beta_{(0,3)} И_3 + \cdots + \beta_{(0,255)} И_{255}}{И'_1 + И'_2 + И'_3 + \cdots + И_{255}}$$

Since,

$$\sum_{m=1}^{M} W_{l_m} = 1$$

Hence,

$$\widehat{\beta}_0 = \beta_{(0,1)} W_1 + \beta_{(0,2)} W_2 + \beta_{(0,3)} И_3 + \cdots + \beta_{(0,255)} W_{255}$$

Step 4: Form Model

MSC	Full Model
AIC_c	$\hat{Y}_i = -0.9709 + 0.0052X_1 + 0.0417X_3 + 0.0277X_4 + 0.0349X_5 - 0184X_6 + 0.0452X_7 - 1.751e^{-6}X_8$
BIC	$\hat{Y}_i = -0.9709 + 0.0052X_1 + 0.0417X_3 + 0.0277X_4 + 0.0349X_5 - 0184X_6 + 0.0452X_7 - 1.751e^{-6}X_8$

Step 5: Elimination of Insignificant Variable

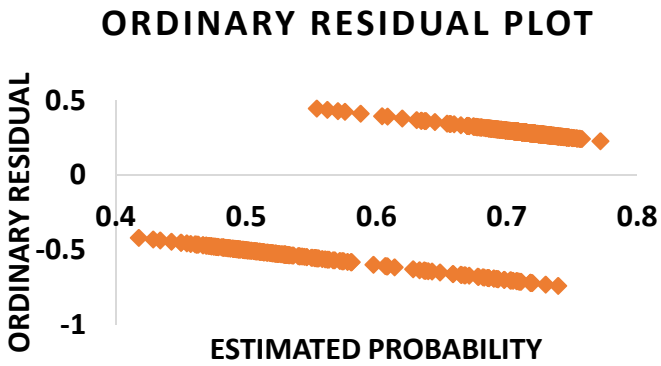
Model Averaging Best Model

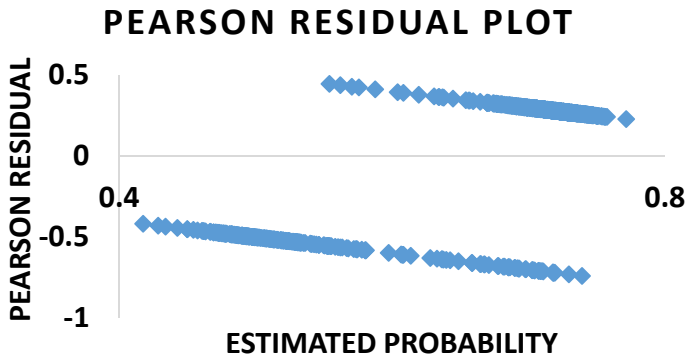
Variable	Elimination 1	Final Model
X_1	<2e-16	<2e-16
X_2	0.0927	
X_3	<2e-16	<2e-16
X_4	<2e-16	<2e-16
X_5	<2e-16	<2e-16
X_6	<2e-16	<2e-16
X_7	<2e-16	<2e-16
X_8	<2e-16	<2e-16

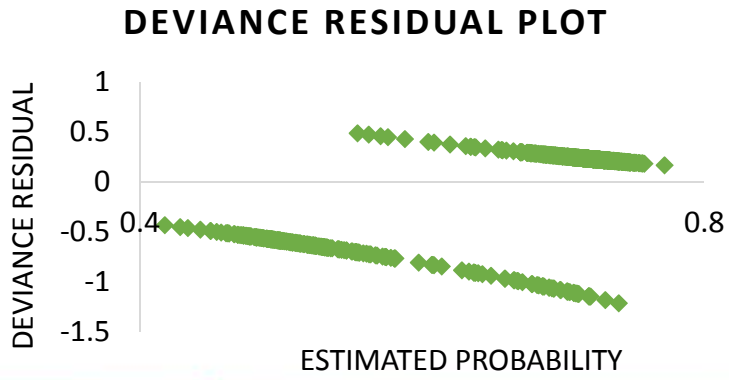
New Model Averaging Best Model

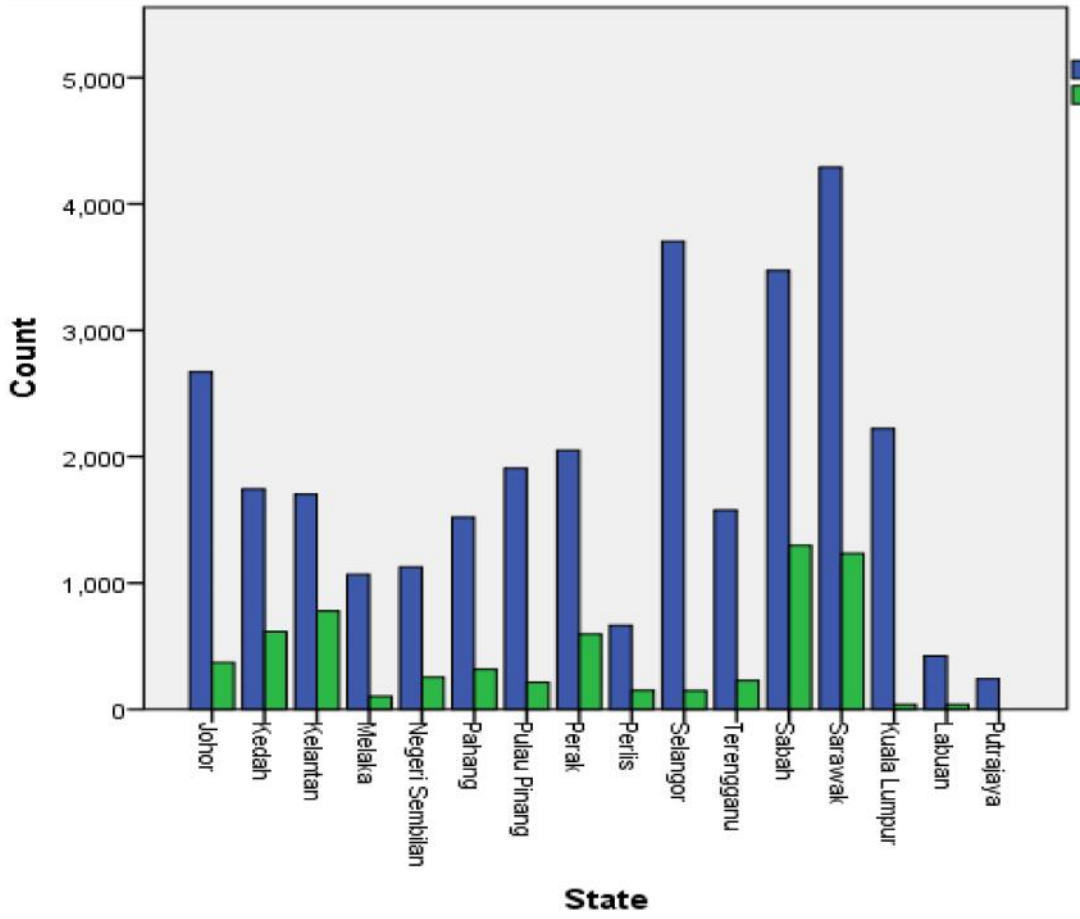
Step 6: Goodness-of-fit Test

- Pearson Chi-Square goodness-of-fit test and Deviance goodness-of-fit test as suggested by (Aisyah *et al.* 2018) were carried out.
- Three residual scatter plot were plotted
- Best model should approximately result in horizontal line with zero intercept.

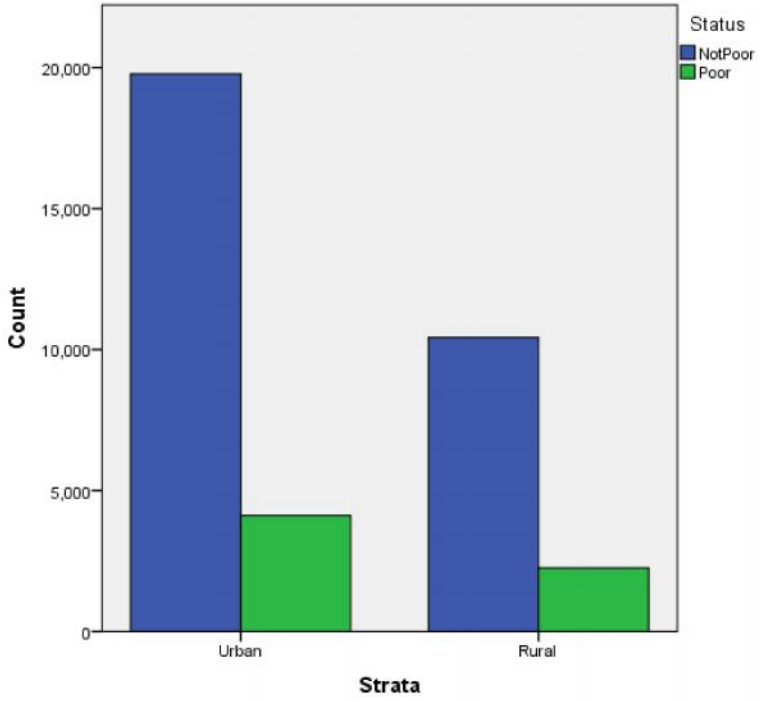




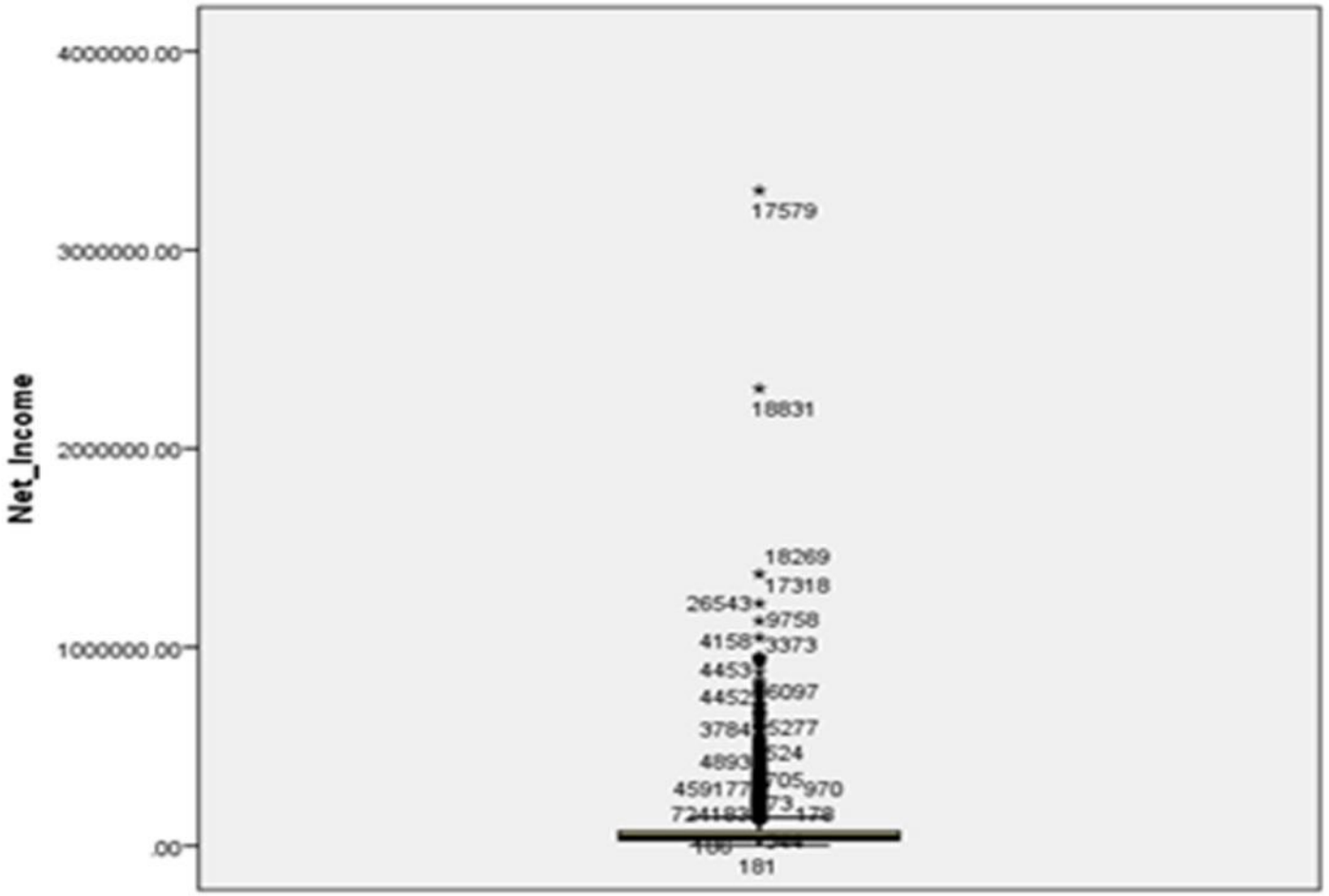




Poverty based on State



Poverty based on Strata


4000000.00
3000000.00
2000000.00
1000000.00
.00
Net_Income
17579
18831
18269
17318
26543
9758
3373
6097
5277
524
970
181

Net Income

RESULT

$$\hat{Y}_i = -0.9709 - (0.0052)X_1 + (0.0417)X_3 + (0.0277)X_4 + (0.0349)X_5 - (0.0184)X_6 + (0.04524)X_7 + (1.7510e^{-6})X_8$$

$$P_i(Y_i = 1) = P_i = \frac{exp^{-0.9709}}{1 + exp_{-0.9709}} = 0.274$$

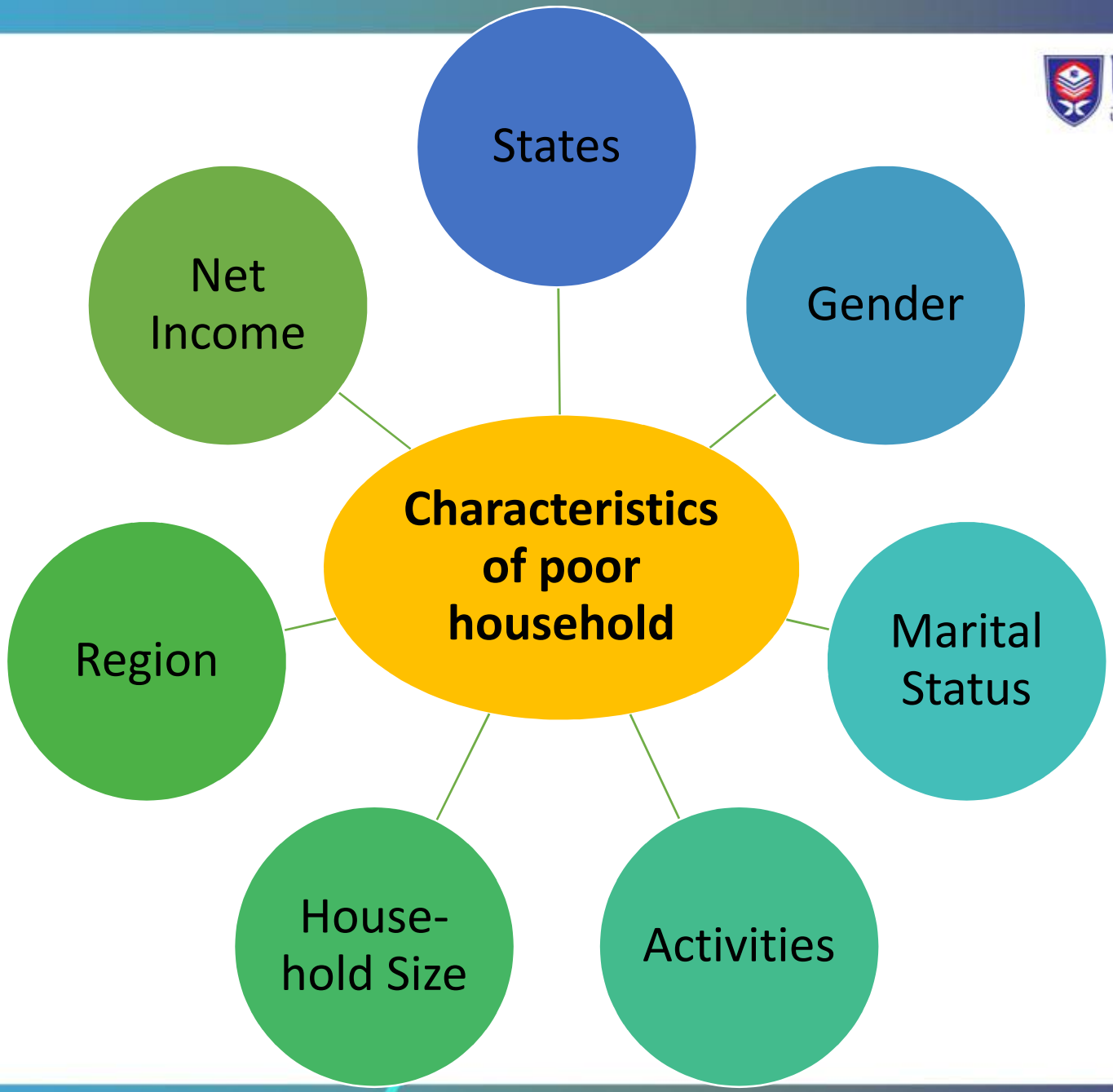
$$P_i(Y_i = 0) = 1 - (0.274) = 0.7253$$

Since X_1 and X_6 have a negative effect on poverty, one unit increase in each covariates will increase the probability of household poverty by 0.0052 and 0.0184 respectively. As for other covariates with positive effect on dependent variable, the probability of poverty will decrease if there is one unit increase in covariates.

CONCLUSION

- Best Model

$$\widehat{Y}_i = -0.9709 - (0.0052)\text{State} + (0.0417)\text{H. Gender} + (0.0277)\text{H. Marital} + (0.0349)\text{H. Activity} + (0.0184)\text{H. Size} + (0.04524)\text{Region} + \left(1.7510e^{-6}\right)\text{N. Inc}$$


States
Net Income
Gender
Characteristics of poor household
Region
Marital Status
House-hold Size
Activities

1. Anyanwu JC, (2014), Marital status, household size and poverty in Nigeria: evidence from the 2009/2010 survey data. *African Development Review*, 26(1) pp. 118-137.

2. Bursac Z. *et al*. (2008). Purposeful Selection of Variables in Logistic Regression: Macro and Simulation Results. *Section on Statistical Computing*, pp. 1886-1891.

3. Claeskens G & Hjort NL, Model Selection and Model Averaging, United Kingdom: University Press, Cambridge, (2008),

4. Hurvich CM & Tsai CL (1989), Regression and time series model selection in small samples. *Biometrika* ,76, pp.297-307.

5. Kutner MH, Nachtsheim CJ, & Neter, J, Applied Linear Regression Models 4th Edition, McGraw-Hill Inc.: Singapore, (2008).

6. Parthiban SG (2018), Poverty measurement revisited from a multidimensional perspective among Universiti Sains Malaysia's B40 poor students. *Malaysian Journal of Society and Space*, 14(4), pp.299-307.

7. Posada D and Buckley TR (2004), Model Selection and Model Averaging in Phylogenetics:advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, 53(5), pp. 793-808.11.

8. Raftery AE (1999), Bayes factors and BIC: Comment on A critique of the Bayesian information criterion for model selection. *Sociol Methods Research* 27, pp. 411–427.12.

9. Saidatulakmal, “Poverty Issues Among Malaysian Elderly”, *Proceeding of the Social Sciences Research*, (2014), pp. 123-132.

10. Schwarz G, (1978), Estimating the dimension of a model *Annals of Statistics* 6 pp 461–464

11. Siwar, Chamhuri, et al. "Urbanization and urban poverty in Malaysia: consequences and vulnerability." Journal of Applied Sciences 16.4 (2016): 154-160.

Thank You

e-mail: khuneswari@uthm.edu.my

Model based on data year 2012

Variable and Description

Variable	Description			
Y	Poverty Level 0: Not Poor 1: Poor			
X_1	State 01: Johor 02: Kedah 03: Kelantan 04: Melaka	05: Negeri Sembilan 06: Pahang 07: Pulau Pinang 08: Perak	09: Perlis 10: Selangor 11: Terengganu 12: Sabah	13: Sarawak 14: Kuala Lumpur 15: Labuan 16: Putrajaya
X_2	Household Age (H.Age) Age of the head of household			
X_3	Household Gender (H. Gen) 1: Male 2: Female			
X_4	Household Marital (H. Mar) 1: Never married 2: Married 3: Widowed	4: Divorced 5: Separated		
X_5	Household Education (H. Edu) Highest level of formal education			
X_6	Household Activity (H. Act) 1: Employer 2: Government employee 3: Private employee 4: Own account worker	5: Unpaid family worker 6: Unemployed 7: Housewife 8: Student	9: Pensioner 10: Others 11: Child not at school	
X_7	Household Size (H. Size) Total number of household member			
X_8	Region 1: Peninsular Malaysia 2: Sabah (including Labuan) 3: Sarawak			
X_9	Net Income (N. Inc) Net total is the total amount of income			

Step 5: Elimination of Insignificant Variable

Model Averaging Best Model

Variables in MA(AIC_c)	P-VALUE	
	Elimination1	Elimination2
constant	< 2e-16	< 2e-16
X_1	3.80E-06	4.90E-07
X_2	8.54E-05	8.32E-05
X_3	< 2e-16	1.44E-08
X_4	2.00E-07	1.59E-07
X_5	< 2e-16	< 2e-16
X_6	< 2e-16	< 2e-16
X_7	< 2e-16	< 2e-16
X_8	0.946	
X_9	< 2e-16	< 2e-16

Variables in MA (BIC)	P-VALUE	
	Elimination1	Elimination2
constant	< 2e-16	< 2e-16
X_1	8.00E-07	6.04E-07
X_2	0.00257	0.00255
X_3	< 2e-16	1.64E-08
X_4	2.00E-07	2.20E-07
X_5	< 2e-16	< 2e-16
X_6	< 2e-16	< 2e-16
X_7	< 2e-16	< 2e-16
X_8	< 2e-16	
X_9	0.98153	< 2e-16

New Model Averaging Best Model

Best Model for all Approach and Accuracy Measures

Method	Full Model	RMSE	MSE	MAE
MA using AIC_c	$\hat{Y}_i = -0.1030 + 0.001952X_1 + 0.000598X_2 + 0.02808X_3 + 0.01716X_4 + 0.0007363X_5 + 0.01717X_6 - 0.0074845X_7 + 1.023e^{-4}X_8 + +6.098e^{-7}X_9$	0.1928**294**	0.08**64072**	0.0371**832**
MA using BIC	$\hat{Y}_i = -0.1022 + 0.001957X_1 + 0.0005739X_2 + 0.02803X_3 + 0.01729X_4 + 0.000736X_5 + 0.01725X_6 - 0.007483X_7 - 7.49e^{-6}X_8 + 6.098e^{-7}X_9$	0.1928**307**	0.08**65210**	0.0371**833**
NMA using AIC_c	$\hat{Y}_i = -0.1029 + 0.001963X_1 + 0.0005984X_2 + 0.02809X_3 + 0.01716X_4 + 0.0007364X_5 + 0.01717X_6 - 0.0078425X_7 + 6.101e^{-7}X_9$	**0.1928272**	**0.0859525**	**0.0371823**
NMA using BIC	$\hat{Y}_i = -0.1022 + 0.001959X_1 + 0.000574X_2 + 0.02803X_3 + 0.01729X_4 + 0.000736X_5 + 0.01725X_6 - 0.007483X_7 + 6.101e^{-7}X_9$	0.1928**274**	0.08**59573**	0.0371**824**

RESULT

BEST MODEL

$$\widehat{Y}_i = -0.1029 + (0.001963)\text{State} + (0.0005984)\text{H. Age} + (0.02809)\text{H. Gender} + (0.01716)\text{H. Marital} + (0.0007364)\text{H. Education} + (0.01717)\text{H. Activity} - (0.0078425)\text{H. Size} + (6.101e^{-7})\text{N. Inc}$$

$$P_i = \frac{exp^{-0.1029}}{1 + exp^{-0.1029}} = 0.4743$$

One-unit increase in State (X1) will decrease the probability of household poverty by 0.001963. Similarly, the probability of poverty also will decrease if there is an increase in variables Age, Gender, Marital, Education, Activity, and Net Income.