KEMENTERIAN EKONOMI
JABATAN PERANGKAAN MALAYSIA

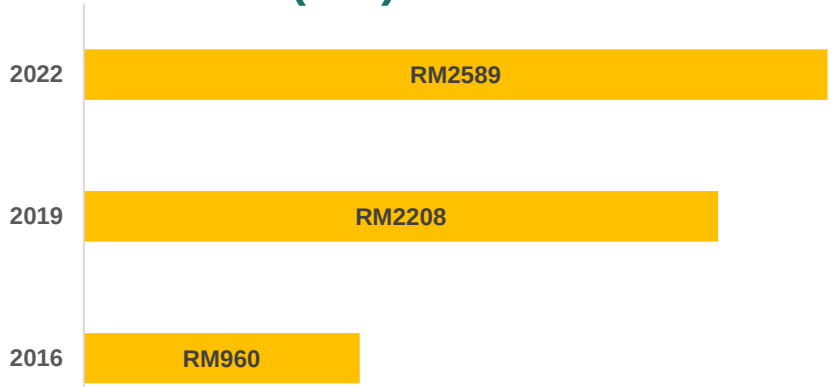
Malaysian Household Poverty Determinant by Using Model Selection & Model Averaging

BY MISS SITI AISYAH MOHD PADZIL

PERSIDANGAN STATISTIK KALI KE-11
"Data dan Kecerdasan Buatan: Memperkasa Masa Depan"
19 September 2024
Dianjurkan oleh:
KEMENTERIAN EKONOMI JABATAN PERANGKAAN MALAYSIA
BANK NEGARA MALAYSIA
MALAYSIA INSTITUTE OF STATISTICS
MyStats 202

INTRODUCTION

Poverty Line Income (PLI) on in Peninsular Malaysia



- ✓ **Poverty incidence had increased from 405.4 thousand in 2019 to 639.8 thousand in the following year (DOSM, 2020).**

- ✓ **Impact of poverty in a country**
- **Effect the Individual life expectancy (Khaled et al., 2020).**
- **Contributes to malnutrition (Faareha et al. 2020).**
- **Poor Education (Darshana et al. 2021).**

- ✓ **This research aims to provide some insight to help policy makers in Malaysia in organizing strategies to eradicate poverty by summarizing the most influential determinants of poverty for year 2019.**

INTRODUCTION

Model Selection	VS	Model Averaging	VS	Hybrid Method Variable Selection Model Averaging

Research Objectives

- **Identify** the **determinants of poverty** by using the 2019 Household Income & Expenditure data provided by DOSM.

- Conducts a **comparative analysis** using three statistical modelling methods: **model selection, model averaging, and variable selection model averaging.**

Methodology

Data:Household Income & Expenditure 2019

Variables	Description	
Y	**Poverty Status**	
	0: Poor	1: Not Poor
X1	**Household Size**	
X2	**State**	
	1: Johor	9: Perlis
	2: Kedah	10: Selangor
	3: Kelantan	11: Terengganu
	4: Melaka	12: Sabah
	5: Negeri Sembilan	13: Sarawak
	6: Pahang	14: W.P. Kuala Lumpur
	7: Pulau Pinang	15: W.P. Labuan
	8: Perak	16: W.P. Putrajaya
X3	**Strata**	
	1: Urban	2: Rural
X4	**Ethnic**	
	1: Bumiputera	3: Indian
	2: Chinese	4: Others
X5	**Gender**	
	1: Male	2: Female
X6	**Age**	
X7	**Highest Certificate**	4: STPM
	1: Degree/Advance	5: SPM/ SPMV
	2: Diploma	6: PMR/SRP
	3: Diploma / certificate	7: No Certificate
X8	**Activity Status**	
	1: Employer	8: Student
	2: Government employee	9: Government pensioner
	3: Private employee	10: Private pensioner
	4: Own account worker	11: Elderly
	5: Unpaid family worker	12: Persons with
	6: Unemployed	15: Others
	7: Housewife	
X9	**Occupation**	
	1: Manager	7: Craft and related trades work
	2: Professional	8: Plants and machine operators
	3: Associate professionals	9: Elementary occupations
	4: Clerical support workers	10: Not classified
	5: Services and sales workers	
	6: Skilled agricultural	
X10	**Working Industry**	
	1: Agriculture, forestry and fishing	11: Financial
	2: Mining and quarrying	12: Professional
	3: Manufacturing	13: Administrative
	4: Electricity supply	14: Public administration
	5: Water supply	15: Education
	6: Construction	16: Human health
	7: Wholesale and retail trade	17: Arts
	8: Transportation and storage	18: Other service activities
	9: Accommodation and food services	19: Household as employers
		20: Organizations
	10: Information and communication	21: Industries not classified

Sample Size: 16, 354

METHODOLOGY

❖ **Logistic Regression Model**

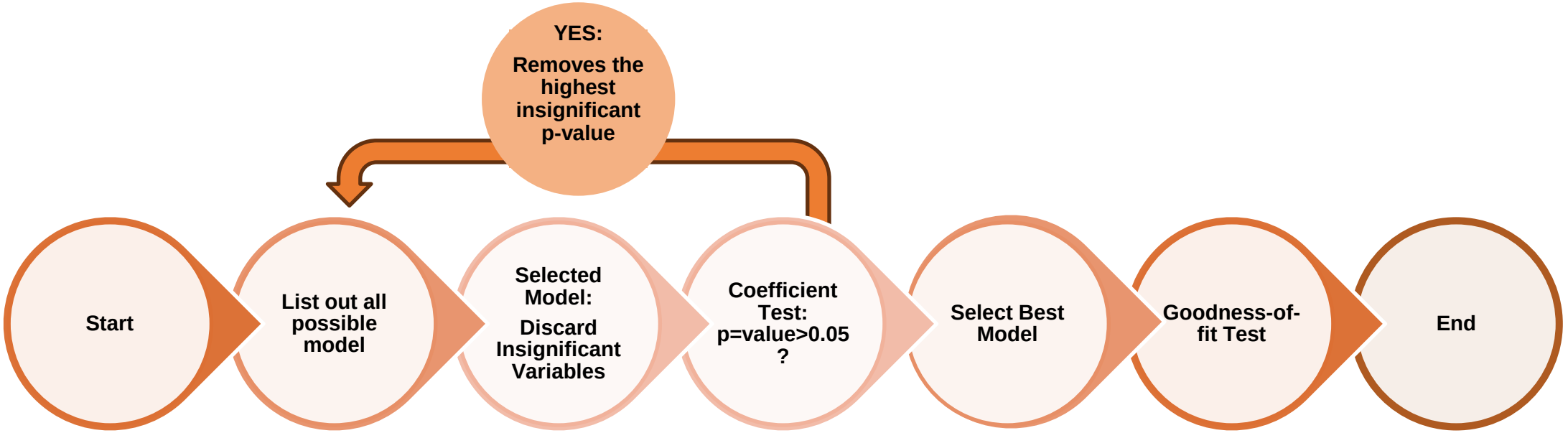
$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_q X_{qi} + u_i$$

❖ **Modelling methods**

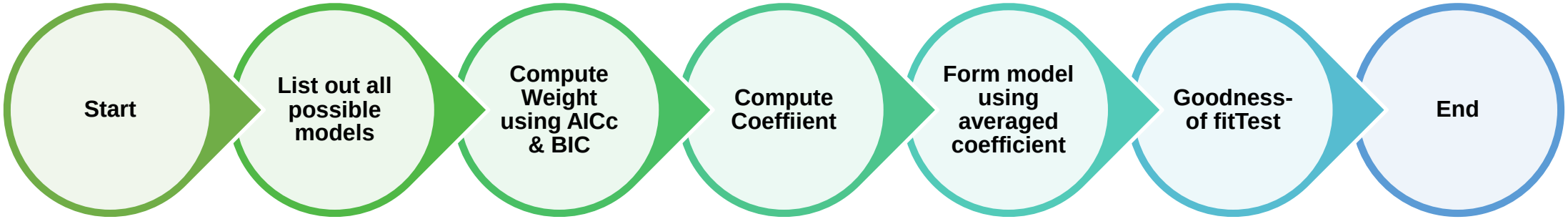
Model Selection	Model Averaging	Variable Selection Model Averaging
• Backward Elimination	• Compare Information Criteria (AICc & BIC)	• Compare Information Criteria (AICc & BIC)

❖ **Brier Score are used for model accuracy comparison**

MODEL SELECTION FLOWS



MODEL AVERAGING FLOWS



Weight Computation

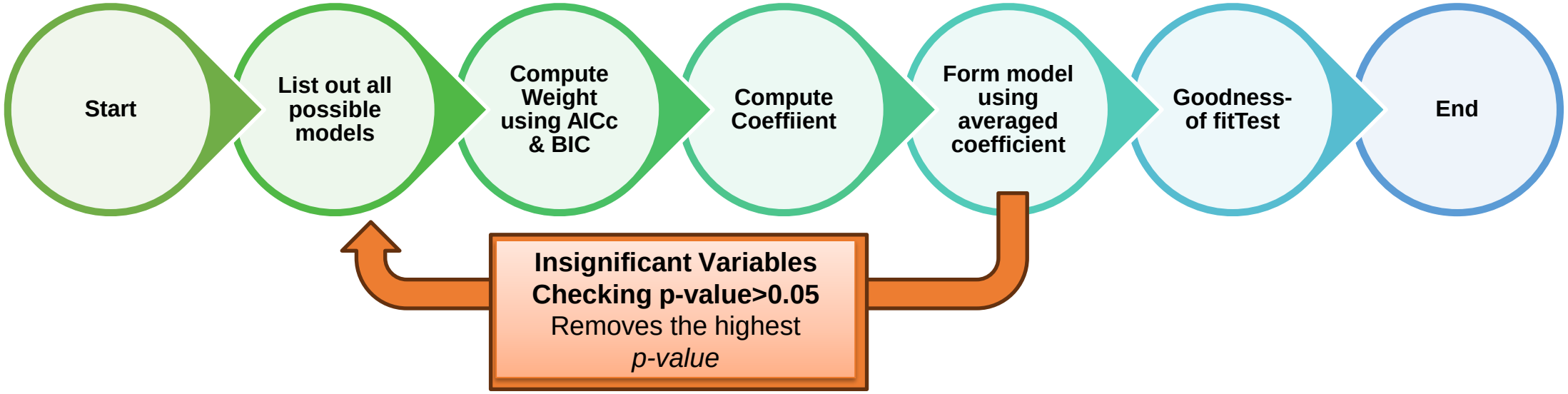
$$W_m = \frac{\exp(\frac{I_m}{2})}{\sum_{m=1}^{M} \exp \frac{I_m}{2}}$$

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

Coeffcient Estimate

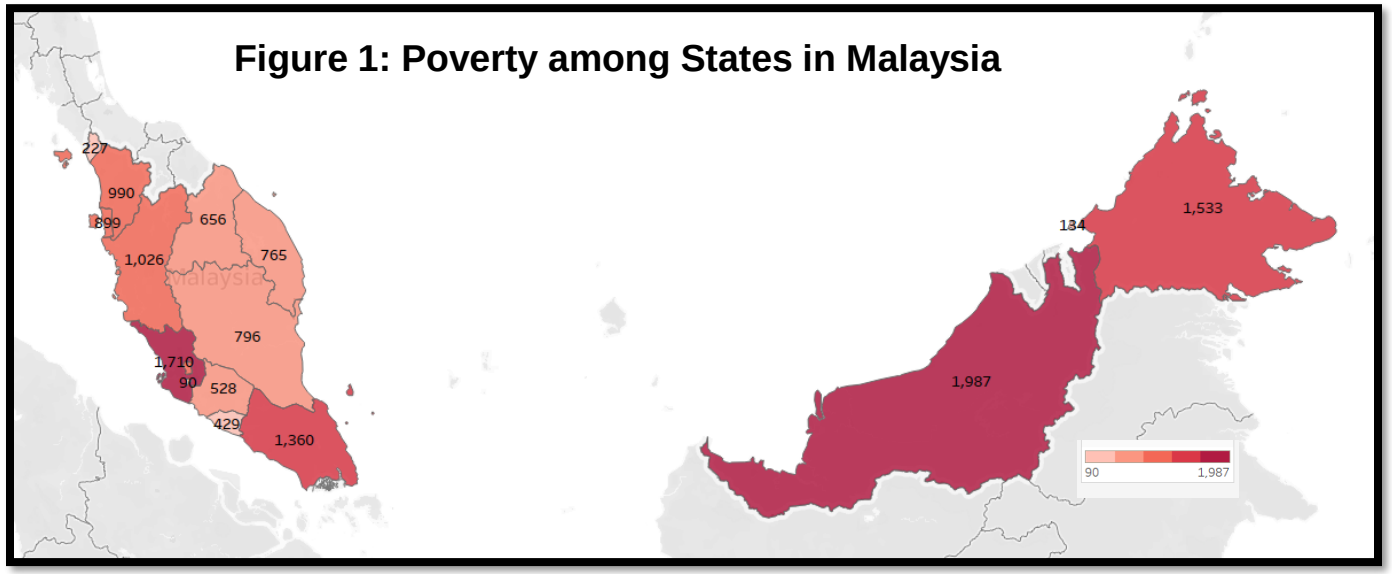
$$\hat{\beta}_p = \sum_{m=1}^{M} w_m \hat{\beta}_{(p,m)}$$

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



VARIABLE SELECTION MODEL AVERAGING FLOWS

Results



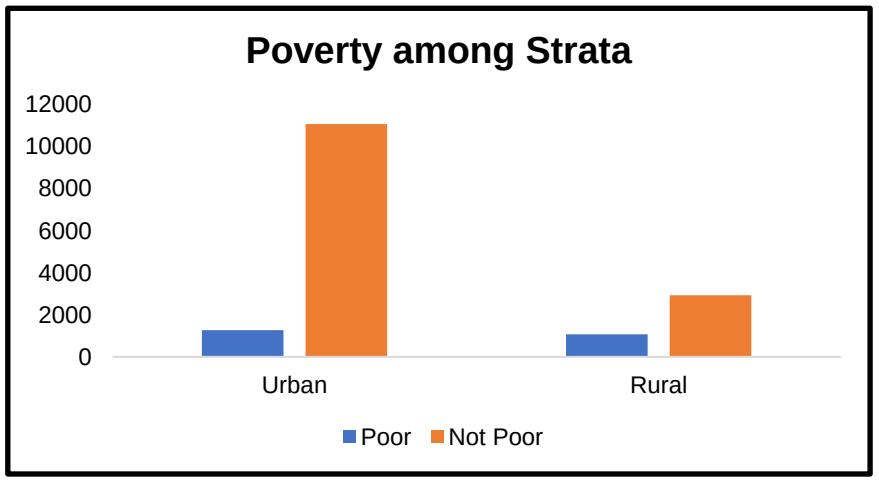


Figure 2: Poverty among Strata in Malaysia

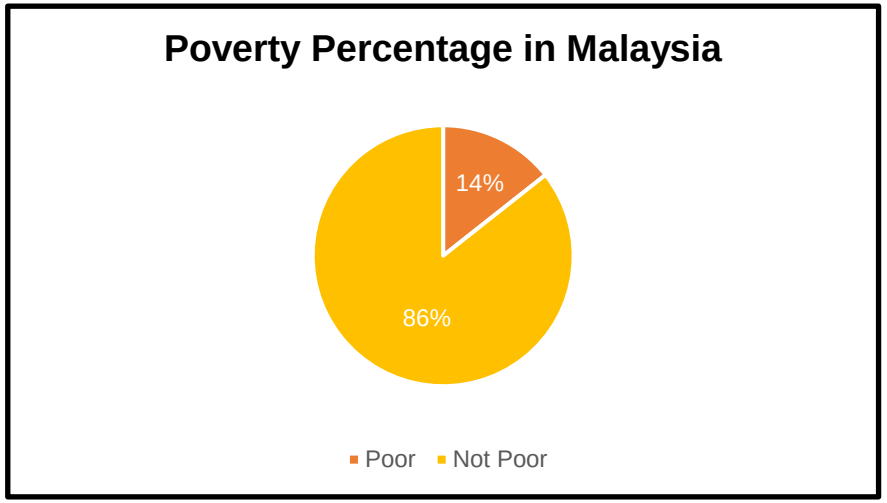
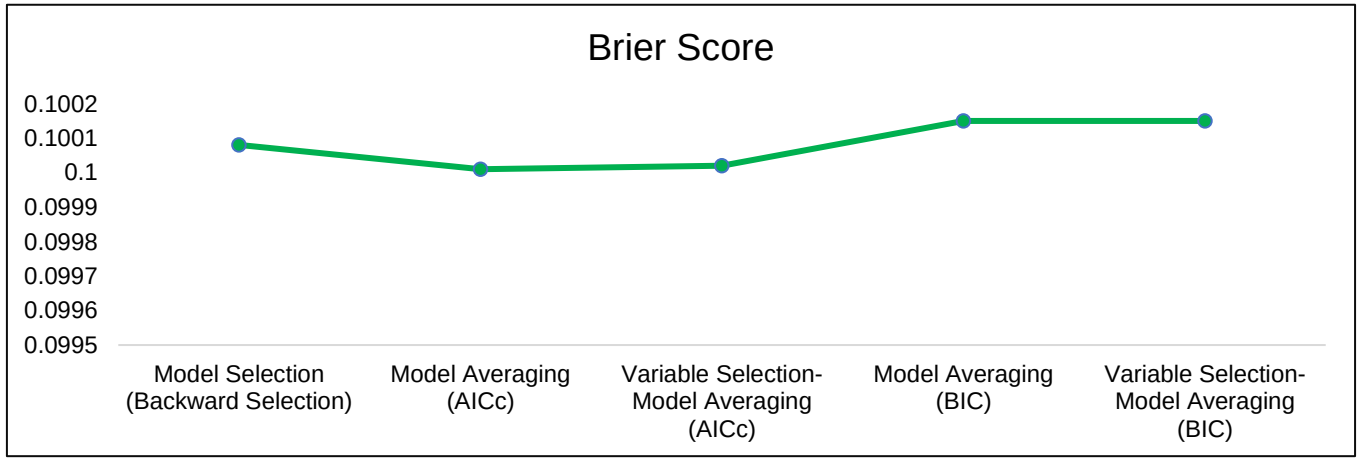


Figure 3: Poverty Percentage in Malaysia

Model Accuracy Comparison

Modelling Methods	Models	Brier Score
Model Selection	$\hat{Y} = -5.15248 - 0.53426\boldsymbol{X1} + 0.88162\boldsymbol{X3} - 0.164513\boldsymbol{X4} + 0.33862\boldsymbol{X5} - 0.01688\boldsymbol{X6} + 0.11\boldsymbol{688}\boldsymbol{X7} + 0.11652\boldsymbol{X8} + 0.57489\boldsymbol{X9}$	0.10008
Model Averaging (AICc)	$\hat{Y} = -5.1589 - 0.5335\boldsymbol{X1} + 0.004676\boldsymbol{X2} + 0.88195\boldsymbol{X3} - 0.16437\boldsymbol{X4} + 0.34405\boldsymbol{X5} - 0.01671\boldsymbol{X6} + 0.11658\boldsymbol{X7} + 0.575567\boldsymbol{X8} + 0.254525\mathbf{X9} + 0.001605\mathbf{X10}$	0.10001
Model Averaging (BIC)	$\hat{Y} = -5.10572 - 0.5308\boldsymbol{X1} + 0.004712\boldsymbol{X2} + 0.8799\boldsymbol{X3} - 0.16862\boldsymbol{X4} + 0.389802\boldsymbol{X5} - 0.01555\boldsymbol{X6} + 0.117049\boldsymbol{X7} + 0.579096\boldsymbol{X8} + 0.255064\mathbf{X9} + 0.001607\mathbf{X10}$	0.10015
Variable Selection-Model Averaging (AICc)	$\hat{Y} = -5.15248 - 0.5338\boldsymbol{X1} + 0.88131\boldsymbol{X3} - 0.16508\boldsymbol{X4} + 0.345582\boldsymbol{X5} - 0.0167\boldsymbol{X6} + 0.11653\boldsymbol{X7} + 0.57548\boldsymbol{X8} + 0.255015\mathbf{X9}$	0.10002
Variable Selection-Model Averaging (BIC)	$\hat{Y} = -5.10525 - 0.5308\boldsymbol{X1} + 0.8799\boldsymbol{X3} - 0.1686\boldsymbol{X4} + 0.38952\boldsymbol{X5} - 0.01555\boldsymbol{X6} + 0.11705\boldsymbol{X7} + 0.5791\boldsymbol{X8} + 0.25508\mathbf{X9}$	0.10015



Poverty Determinant based on Final Model

$$\widehat{Y} = -5.15248 - 0.5338\boldsymbol{H.Size} + 0.88131\boldsymbol{Strata} - 0.16508\boldsymbol{Ethnic} + 0.345582\boldsymbol{Gender} - 0.0167\boldsymbol{Age} + 0.11653\boldsymbol{Certificate} + 0.57548\boldsymbol{Activity} + 0.255015\mathbf{Occupation}$$

- Household Size
- Strata
- Ethnic
- Head of household gender
- Head of household age
- Head of household activity status
- Certificate
- Head of household type of occupation

Poverty Determinant

- State
- Working Industries

Omitted Variable

Conclusion

- Slight difference between Model Averaging and Variable Selection Model Averaging accuracy. If the decimal values are rounded to one decimal place, the error would be identical.
- The best model can be determined by the researcher's objectives of study. If the aim is to point out the most contributing covariates, Variable Selection Model Averaging using AICc is comparable as a standard method.
- Finally, the determinants of poverty in year 2019 are household size, strata, ethnic, head of household gender, head of household age, head of household activity status, certificate and head of household type of occupation.

References:

1. Burnham and Anderson. 2002. Model Selection and Multimodel Inference:A practical Information-theoretic Approach. 2nd Ed. NewYork: Springer-Verlag.
2. Darshana Darmalinggam , Maniam Kaliannan and Mageswari Dorasamy. 2021. Proactive measures to eradicate Malaysia's poverty in IR4.0 era: a shared prosperity vision. F1000 Research. 28;10:1094. doi: 10.12688/f1000research.73330.2. PMID: 35237432; PMCID: PMC8790706.
3. Department of Statistic Malaysia. 2020. Household Income & Basic Amenities Survey Report 2019. Reference: https://v1.dosm.gov.my/v1/index.php?r=column/ctheme ByCat&cat=120&bulid=TUTmRhQ1N5TUxHVWN0T2VjbXJYZz09&menu_id=amVoWU54UTl0a21 NWmdhMjFMMWcyZz09.
4. Economic Planning Unit, Prime Minister Department Malaysia. 2017. Malaysia Succes Story in Poverty Eradication.
5. Faareha Siddiqui, Rehana Salam, Zohra Lassi and Jai Das. 2020. The Intertwined Relationship Between Malnutrition and Poverty. Front Public Health. 28;8:453. DOI: 10.3389/fpubh.2020.00453
6. Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2008). Applied Linear Regression Models. 4th edition. Singapore: McGraw-Hill Inc.
7. Khuneswari Gopal Pillay, Siti Aisyah Mohd Padzil & Noraini Abdullah. 2018. Model Selection and Model Averaging on Mortality of Upper Gastrointestinal Bleed Patients IOSR Journal of Dental and Medical Sciences: 17, 68–78. https://doi.org/10.9790/0853-1711086878
8. Khaled Thafran, Makmur Tumin and Ahmad Farid Osman. 2020. Poverty, Income and Unemployment as Determinant of Life Expectency: Empirical Evidence from Panel Data of Thirteen Malaysian States. Iranian Journal of Pubic Health
9. Mu Y, See I, Edwards JR. 2019. Bayesian model averaging: improved variable selection for matched case-control studies. Epidemiol Biostat Public Health. 16(2):13048. DOI: 10.2427/13048. PMID: 31772926; PMCID: PMC6879006.
10. Micheal Schomaker and Christian Heumann. 2014. Model selection and model averaging after multiple imputation. Computational Statistics and Data Analysis (71)758-770.
11. Percetakan Nasional Malaysia Berhad. 2020. Shared Prosperity Vision 2030. Restructuring the priorities of Malaysia Development. Saidatulakmal. 2014. Poverty Issues Among Malaysian Elderly. Proceeding of the Social Sciences Research: 123-132
12. Siti Aisyah Mohd Padzil, Khuneswari Gopal Pillay, Mohd Saifullah Rusiman, & Rohayu Mohd Salleh. 2019. New model averaging approach in predicting mortality rate of intensive care unit patients. Journal of Physics: Conference Series, 1366(1). DOI: https://doi.org/10.1088/1742-6596/1366/1/012123
13. Steyerberg, Harrell, Borsboom et al. 2001. Internal Validation of predictive models: Efficiency of some procedures for logistic regression analysis. Journal of Clinical Epidemiology, 54, pp. 774-781.
14. Zainodin H J and Khuneswari Gopal Pillay. 2010. Model-Building Approach in Multiple Binary Logit Model using Coronary Heart Disease. Malaysian Journal of Mathematical Sciences vol 4) pp 107-133

Thank You

UTHM
Universiti Tun Hussein Onn Malaysia

PERSIDANGAN STATISTIK
KALI KE-11
"Data dan Kecerdasan Buatan: Memperkasa Masa Depan"

19 September 2024

Dianjurkan oleh:

KEMENTERIAN EKONOMI
JABATAN PERANGKAAN MALAYSIA

BANK NEGARA MALAYSIA
CENTRAL BANK OF MALAYSIA

MALAYSIA INSTITUTE
OF STATISTICS

MyStats 202