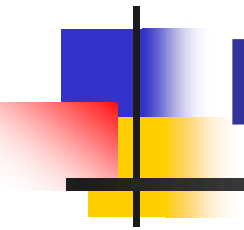
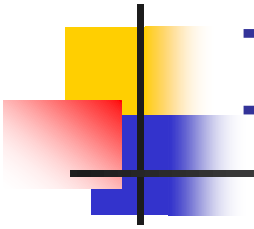


Logistic regression

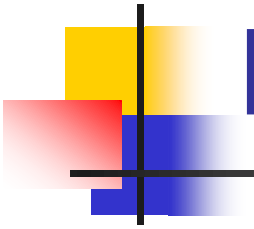
Assignment #1



Introduction

Socio-economic variables are very often categorical, rather than interval scale

logistic regression may be thought of as an approach that is similar to that of multiple linear regression, but takes into account the fact that the dependent variable is categorical.



Definition of Logistic regression

Determines the impact of multiple independent variables presented together to predict membership of one or other of the two dependent variable categories

Assumptions of logistic regression

- Logistic regression does not assume a linear relationship

- The dependent variable must be a dichotomy (2 categories)

- The independent variables need not be interval, nor normally distributed, nor linearly related, nor of equal variance within each group

The Logit Transformation

- linear probability model → $Y_i = BX + \varepsilon_i$
- the ratio of the probability → $odds_i = \frac{P_i}{1 - p_i}$
- the logit or log-odds → $= log \frac{P_i}{1 - p_i}$

The logistic regression equation

$$p = \frac{exp^{(a+b_1x_1+b_2x_2+b_3x_3\ldots)}}{1+exp^{(a+b_1x_1+b_2x_2+b_3x_3\ldots)}}$$

Where:

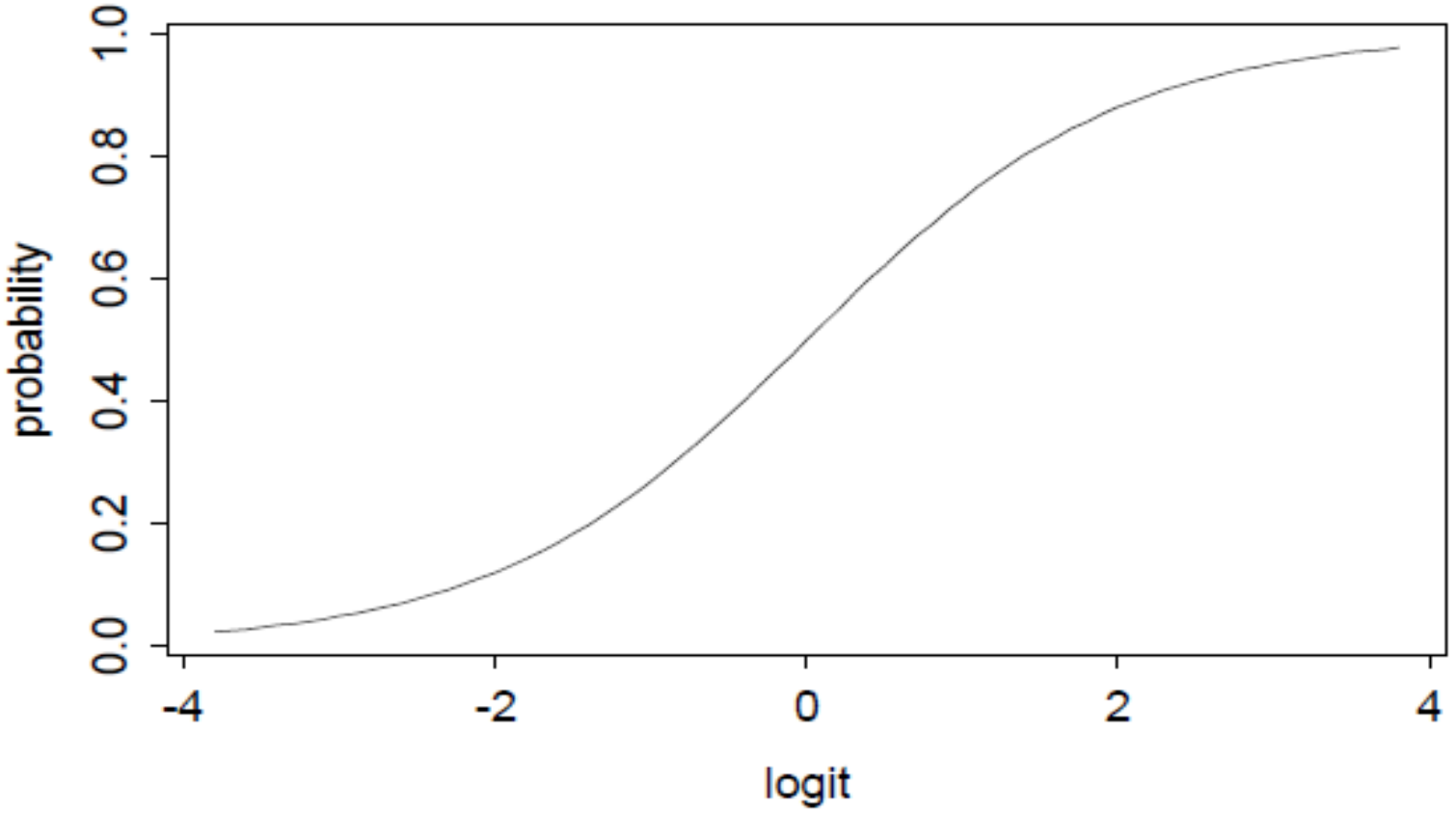
p = the probability that a case is in a particular category,

exp = the base of natural logarithms (approx 2.72),

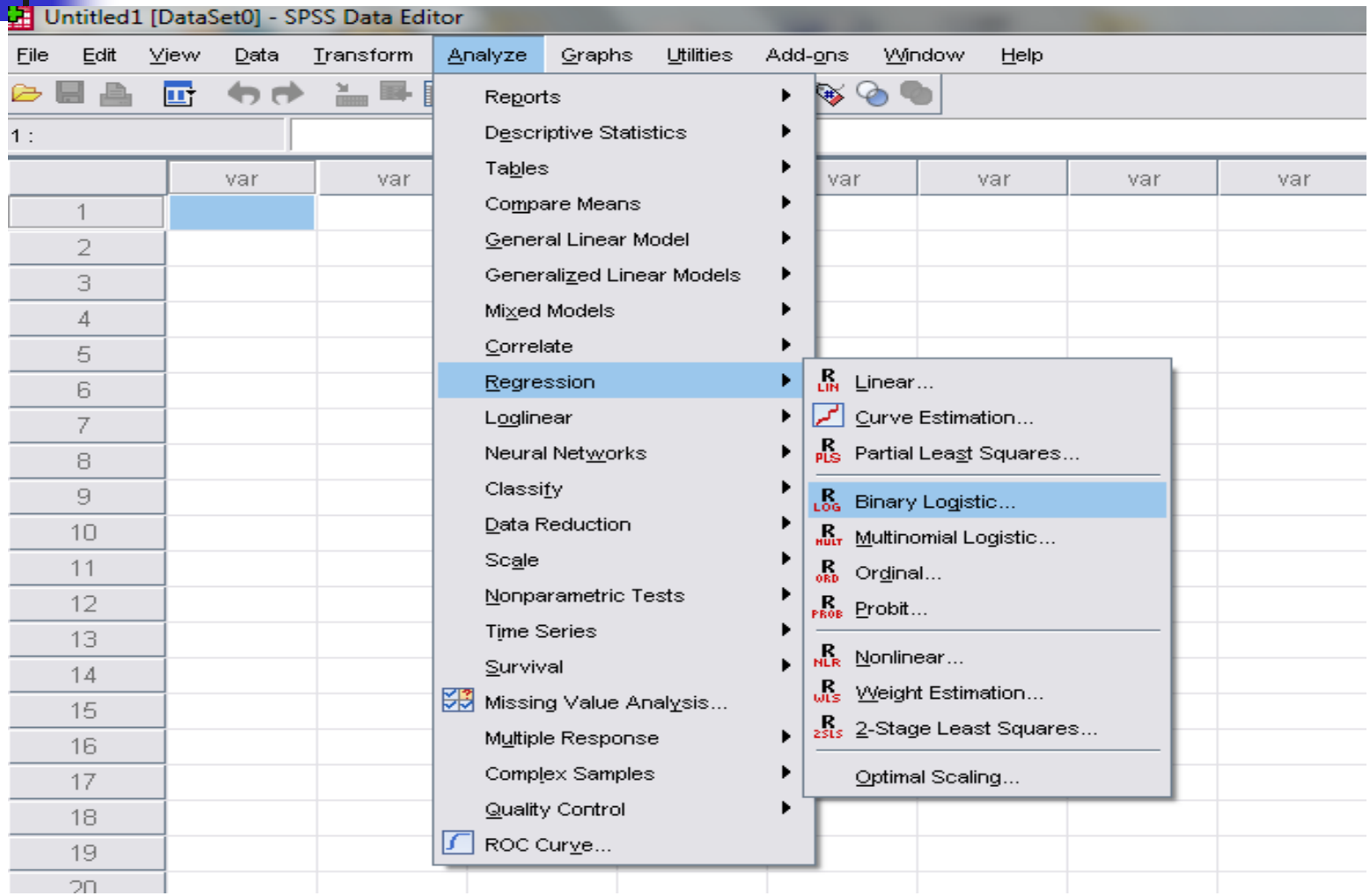
a = the constant of the equation and,

b = the coefficient of the predictor variables

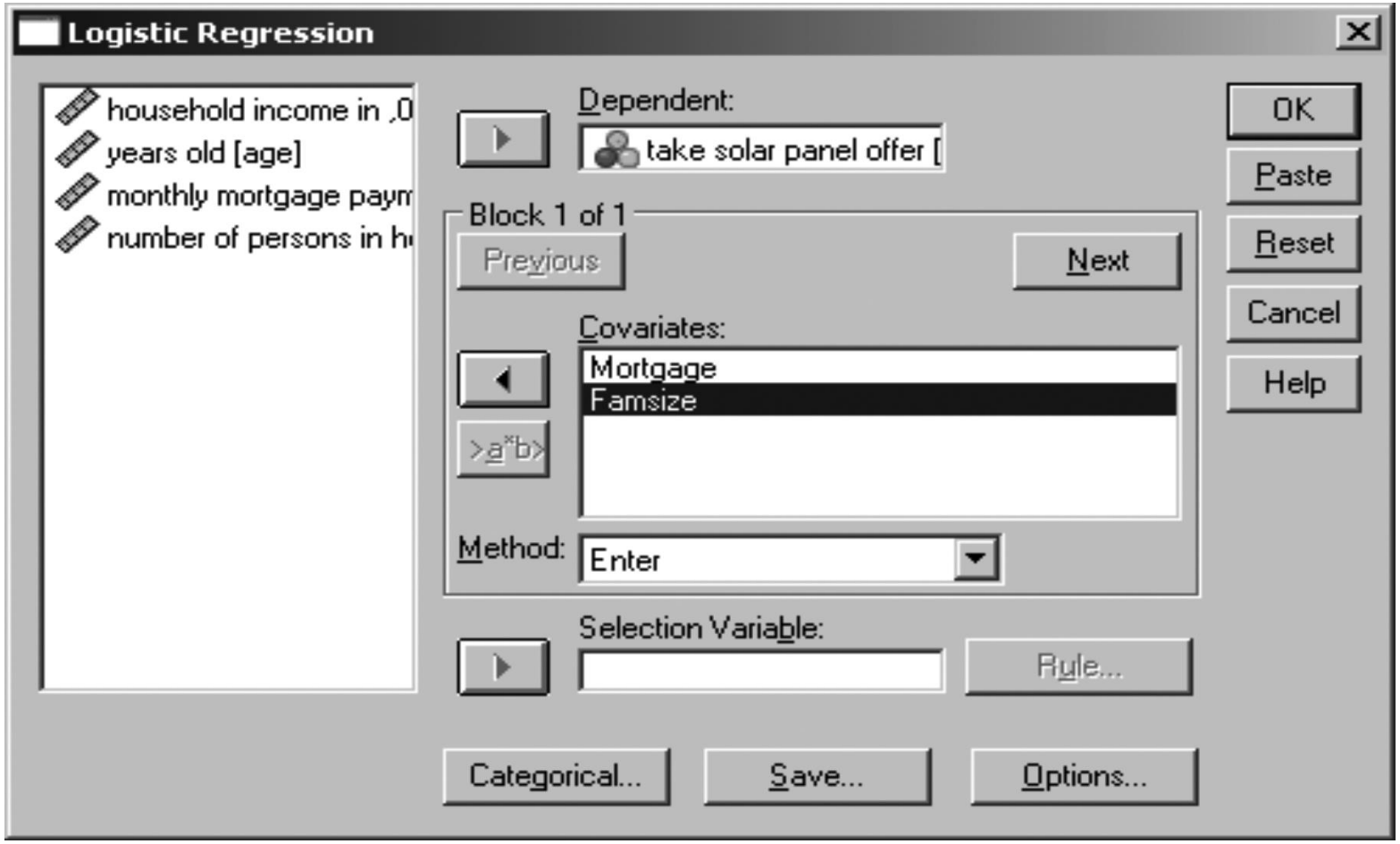
Logistic Curve



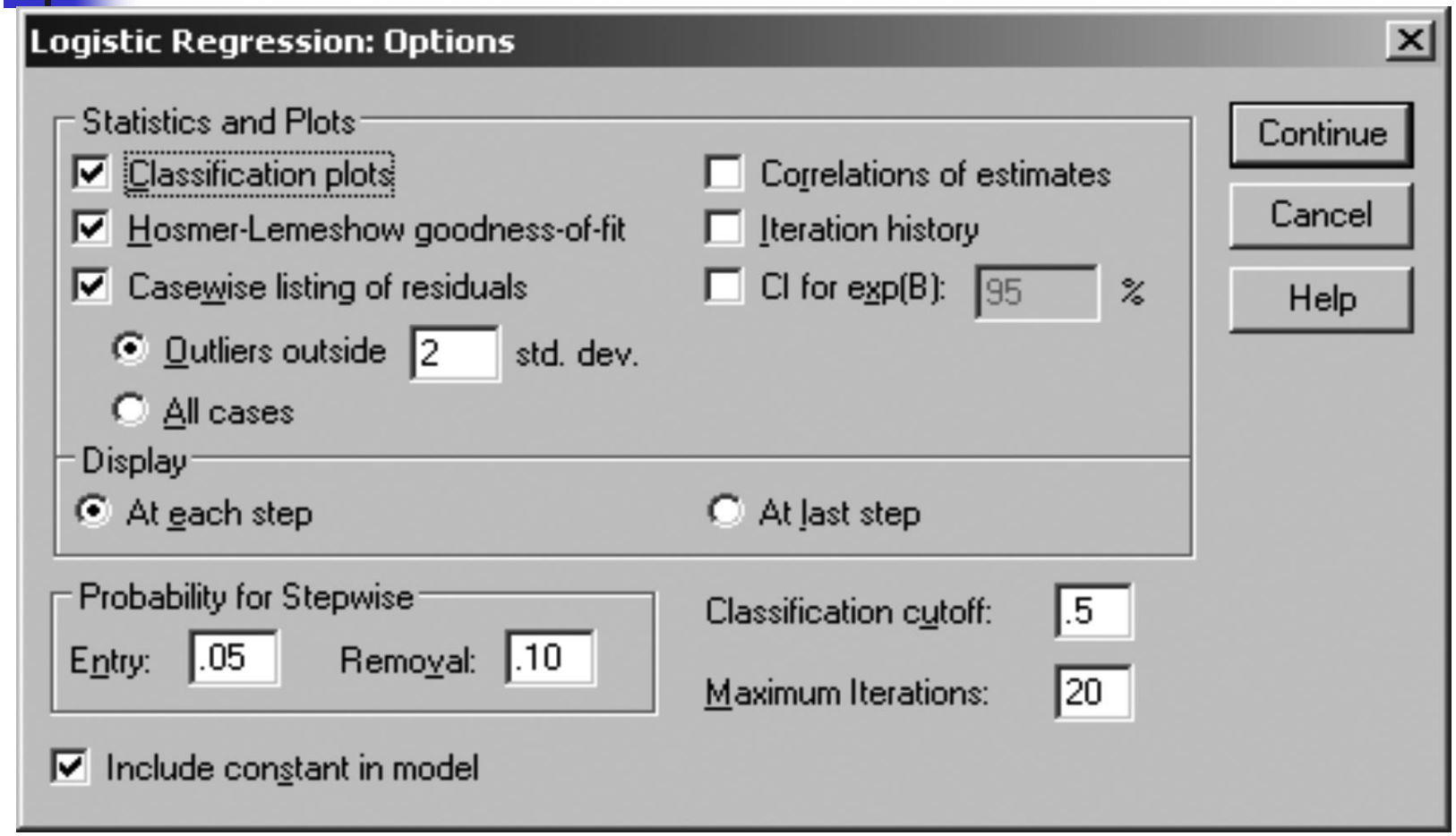
Computer Application



Computer Application



Computer Application



Interpretation of printout

Table (1.1) Classification table

Classification Table[a,b]

Observed			Predicted		
			Take solar panel offer		
			Decline offer	Take offer	Percentage correct
Step 0	take solar panel offer	decline offer	0	14	.0
		take offer	0	16	100.0
	Overall Percentage				53.3

[a] Constant is included in the model.

[b] The cut value is .500.

Interpretation of printout

Table 1.3 Variables not in the equation table

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Famsize	14.632	1	.000
		Mortgage	6.520	1	.011
	Overall Statistics		15.085	2	.001

Interpretation of printout

Table 1.5 Model Summary

Model Summary

Step	−2 Log likelihood	Cox & Snell R square	Nagelkerke R square
1	17.359[a]	.552	.737

[a] Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Interpretation of printout

Table 1.8 Classification table

Classification Table[a]

			Predicted		
			Take solar panel offer		
	Observed		Decline offer	Take offer	Percentage correct
Step 1	take solar panel offer	decline offer	13	1	92.9
		take offer	2	14	87.5
	Overall Percentage				90.0

[a] The cut value is .500.

Interpretation of printout

Table 1.9 Variables in the equation

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Famsize	2.399	.962	6.215	1	.013	11.007
	Mortgage	.005	.003	3.176	1	.075	1.005
	Constant	−18.627	8.654	4.633	1	.031	.000

$$\text{Probability of a case} = \frac{e\{(2.399 \times \text{family size}) + (.005 \times \text{mortgage}) - 18.627\}}{1 + e\{(2.399 \times \text{family size}) + (.005 \times \text{mortgage}) - 18.627\}}$$