

HOMework UNIT1

question 1

Reporter: Xiumei Wang
Instructor: Wolfgang Härdle



Year	RAM	increase
1988	0.002	
1991	0.004	1
1996	0.5	124
2000	1	1
2003	2	1
2007	8	3
2014	16	1

LOGISTIC REGRESSION

With logistic regression, the response variable is an indicator of some characteristic, that is, a 0/1 variable.

Logistic regression is used to determine whether other measurements are related to the presence of some characteristic. For example, whether certain blood measures are predictive of having a disease.

Logistic regression equation does not directly predict the probability that the indicator is equal to 1. It predicts the log odds that an observation will have an indicator equal to 1.

The model logistic regression model is that

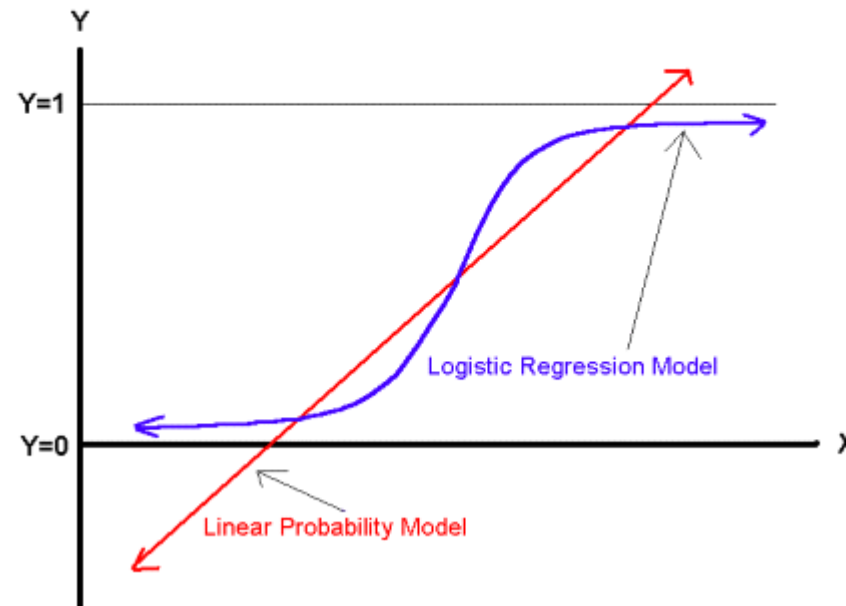
$$P = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}$$

Because logistic regression predicts probabilities, rather than just classes, we can fit it using likelihood.

$$L(\beta_0, \beta) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

The null hypothesis underlying the overall model states that all β s equal zero.

Comparing the LP and Logit Models



We can use the *glm* (generalized linear model) function in R to estimate a logistic regression model using.