



Statistical Corner

Chun-Fan LEE 李俊帆¹, Wilfred Hing-Sang WONG 黃慶生², Daniel Yee-Dak FONG 方以德¹

¹Department of Nursing Studies; ²Department of Paediatrics & Adolescent Medicine, The University of Hong Kong, Hong Kong

Question 1:

What is a logistic regression?

Answer :

It is a type of regression analysis in which the outcome or dependent variable is binary. The factors or independent variables may however be either categorical or quantitative. Effect of a factor is expressed as an odds ratio with 1 indicates null effect.

Question 2:

How to perform logistic regression in SPSS?

(Original data could be obtained from the website of the Society)

Answer:

The use of logistic regression in SPSS is illustrated in a dataset involving 380 patients with prostate cancer (Source: ftp://ftp.wiley.com/public/sci_tech_med/logistic, SPSS format: <http://www.hkspr.org>). To examine the effects of digital rectal examination result (DPROS, categorical with 4 levels: no/left unilobar/right unilobar/bilobar) and tumor volume (VOL, continuous in cm³) on whether tumor has penetrated the prostatic capsule (CAPSULE), we may use "Analyze>Regression>Binary Logistic".

A window will open up as shown in Figure 1. The variable CAPSULE should be put under "Dependent" while DPROS and VOL should be put under "Covariates". By default, SPSS treats all variables put under "Covariates" as continuous. As DPROS is categorical, we need to define it using the "Categorical" button. Also, we need to ask for the confidence intervals using the "Options" button.

Question 3:

How to interpret the results?

Answer:

Figure 2 shows two tables from typical logistic regression in SPSS. It is important to first look at the table "Dependent Variable Encoding". The level of original value that corresponds to the internal value of 1 is what the logistic regression predicts. That is, the logistic regression in the example examines the factor effects of tumor penetration. The second table "Categorical Variables Codings" is only provided when there are categorical factors. Its use is described later when we interpret the factor effects. The last table presents the odds ratios of various factors under the column "Exp(B)" with the corresponding p-values under column "Sig.". The 95% confidence intervals are also what we need to report.

In our example, both VOL ($p=0.042$) and DPROS ($p<0.001$) are significant factors of CAPSULE. In particular, 1 cm³ increase in tumor volume would reduce the risk of penetration by 1.3% ($=1 - 0.987$). Interpreting the effects of the categorical factor DPROS needs to have a reference level defined from which the other levels are compared. For example, patients with no nodule were at lower risk of tumor penetration than patients with bilobar nodule (odds ratio=0.13, $p<0.001$). Note that we are reading on the row "DPROS(1)" which, from table "Categorical Variables Codings", indicates the level of no nodule.

**Question 4:**

How do we know a logistic regression analysis is appropriate?

Answer:

One method to assess the goodness-of-fit of the logistic model is by using the Hosmer-Lemeshow Test (Figure 3). A significant result indicates that the model does not fit well. In this example the model seems to be reasonable with $p = 0.128$.

In case lack of fit for the model, we may consider transforming the covariates or look for any outliers by examining the model residuals. If these cannot remedy the model, alternatives such as complementary log-log model or probit model may be considered.

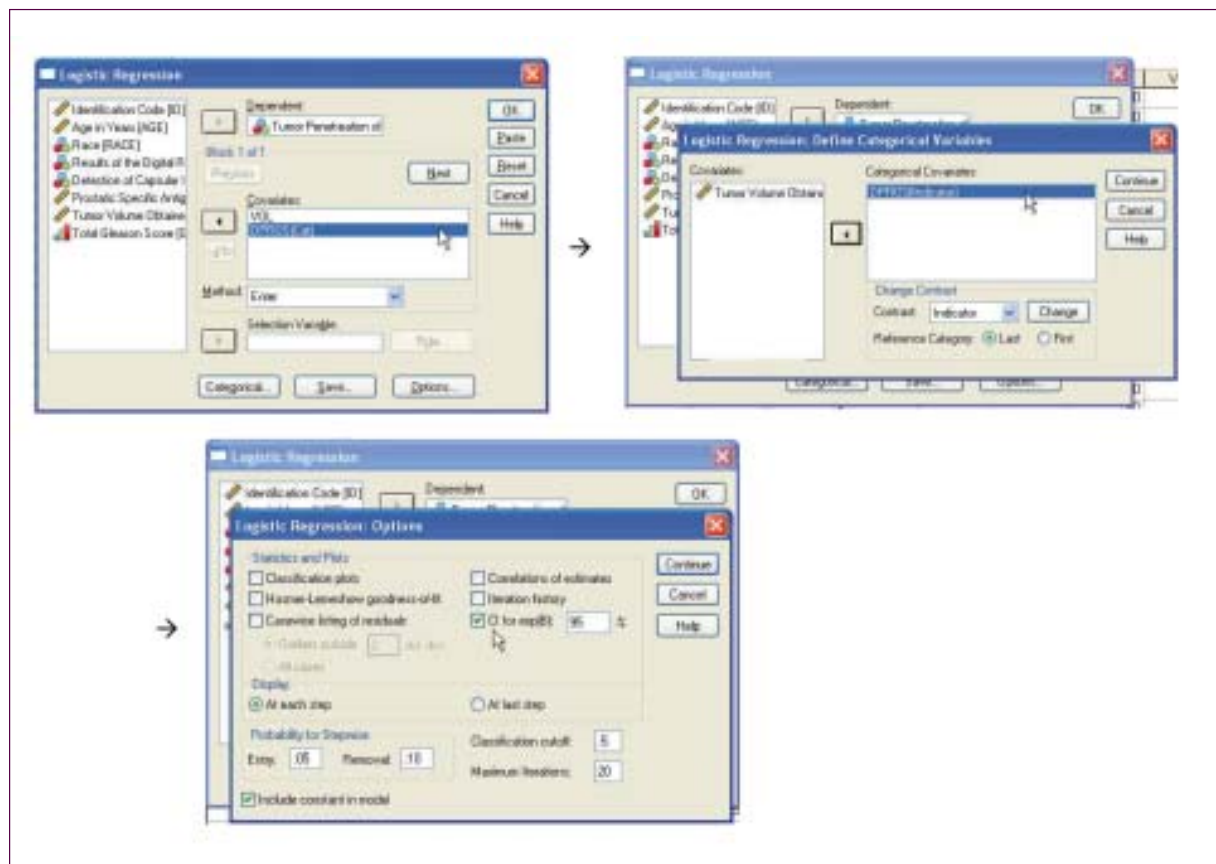


Figure 1. Logistic regression in SPSS

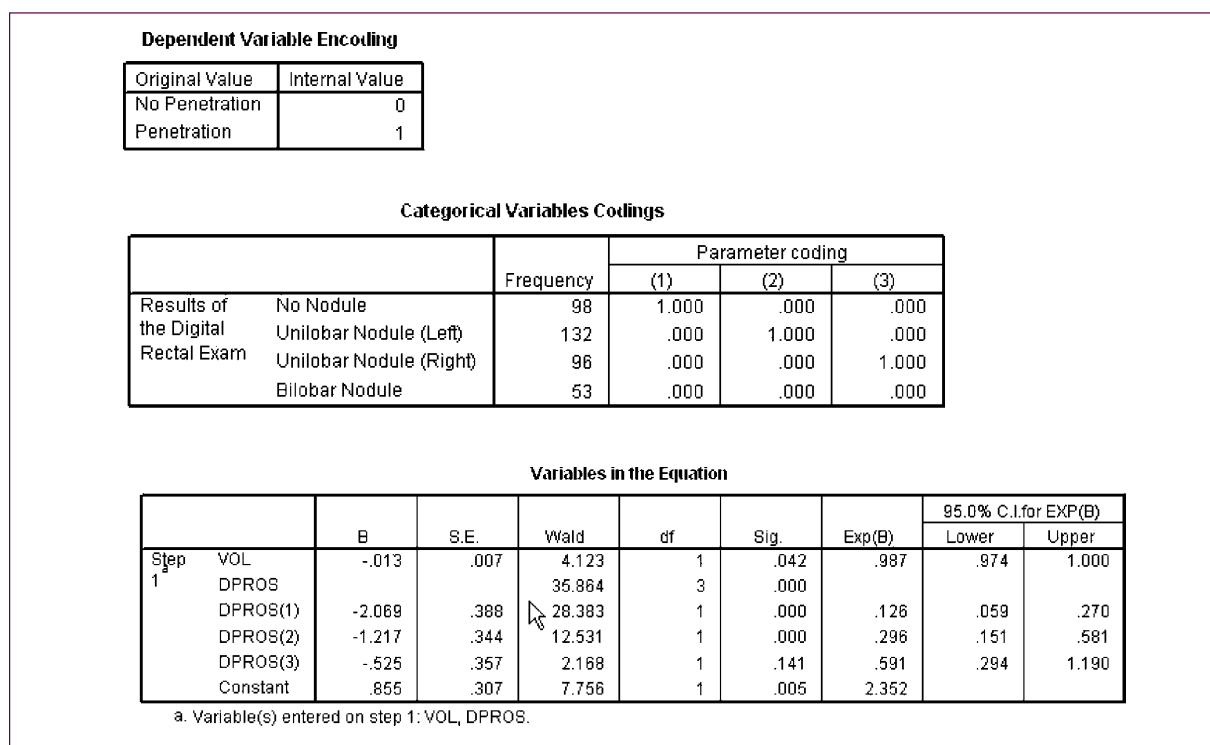


Figure 2. Results of a logistic regression

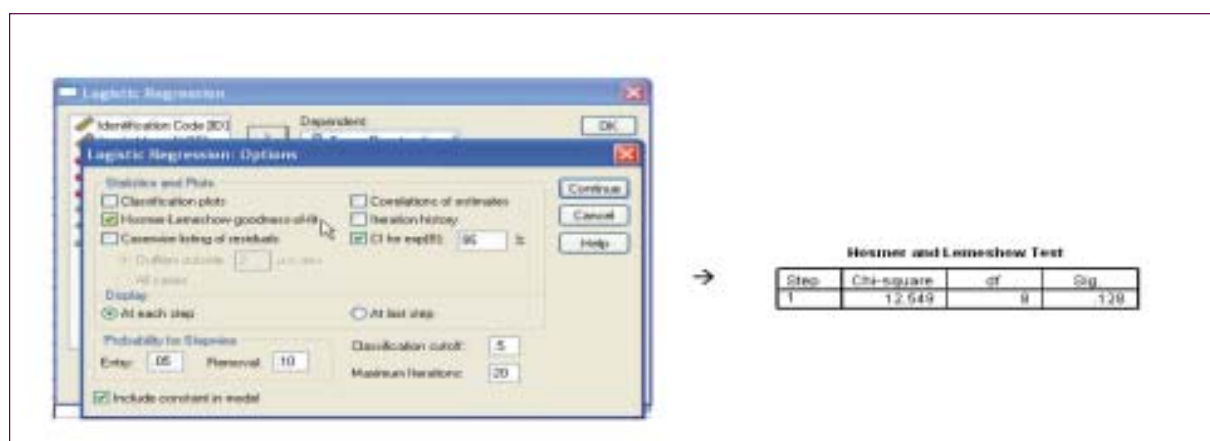


Figure 3. Goodness-of-fit test of a logistic regression model