Logistic Regression Series Part 1: Simple Logistic Regression

Zin Htway, Ph.D., MBA, CT (ASCP, IAC)
Zin Htway, Ph.D., MBA, CT (ASCP, IAC)
Introduction to Simple Logistic Regression

- Logistic regression, also called a logit model, is a test statistic used to model dichotomous outcome variables (in which there are only two possible outcomes) from a set of independent variables which can be discrete and/or continuous.

- Simple logistic regression is limited to a single independent variable and a single dichotomous dependent variable.
How Logistic Regression differs from Linear Regression

• No need for a linear relationship between the dependent and independent variables.
• The independent variables need not to be multivariate normal. – this is not a concern for simple logistic regression.
• Homoscedasticity is not a concern (Homoscedasticity refers to the assumption that that the dependent variable exhibits similar amounts of variance across the range of values for an independent variable).
• Both ordinal and nominal data can be independent variables in logistic regression.
Statistical Assumptions for Logistic Regression

- Binary logistic regression requires the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal.
- Since logistic regression assumes that the function $P(Y = 1)$ is the probability of the event occurring, it is necessary that the dependent variable is coded accordingly.
- The model should be fitted correctly. All meaningful variables should be included. A good approach to ensure this is to use a stepwise method to estimate the logistic regression.
- The error terms need to be independent. Logistic regression requires each observation to be independent.
- Logistic regression assumes linearity of independent variables and log odds.
- Logistic regression requires quite large sample sizes. Multiple logistic regression needs at least 10 cases per independent variable, some statisticians recommend at least 30 cases for each parameter to be estimated.
Review of the foundation basics

- Probability = \( P(\text{event}) = (\text{outcome of interest}) / (\text{all possible outcomes}) \)
- Odds = \( P(\text{event occurring}) / P(\text{event not occurring}) = p / (1 - p) \)
- Maximum value of \( P(\text{event occurring}) = 1.0 = p \)
- Value of \( P(\text{event not occurring}) = (1 - p) \)
- Odds Ratio (OR) = a ratio of two odds = \( (odd_{1}) / (odd_{2}) = [(p_{1} / 1 - p_{1}) / (p_{0} / 1 - p_{0})] \)
Review of Linear Regression

\[ Y = \beta_0 + \beta_1 X \]
Probability in Logistic Regression

$$\pi = \frac{1}{1 + e^{-(\beta_0 + \beta_1X)}}$$
Taking the logit of $\pi$

\[
\text{logit}(\pi) = \log\left(\frac{\pi}{1 - \pi}\right) = (\beta_0 + \beta_1 X)
\]
SPSS for Simple Logistic Regression
Analyze > Regression > Binary Logistic
SPSS for Simple Logistic Regression
Move [DP > Dependent], [IV > Covariates]
SPSS for Simple Logistic Regression
Options > Select: CI for exp(B) 95%
### SPSS Simple Logistic Regression Output using fictitious data set

<table>
<thead>
<tr>
<th>Step 1a</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
<th>95% C.I. EXP(B) Lower</th>
<th>95% C.I. EXP(B) Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV2</td>
<td>0.296</td>
<td>0.036</td>
<td>67.892</td>
<td>1</td>
<td>0.000</td>
<td>1.345</td>
<td>1.254</td>
<td>1.443</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.503</td>
<td>0.534</td>
<td>21.948</td>
<td>1</td>
<td>0.000</td>
<td>0.082</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

What this means is; when there is a one unit increase in the predictor variable this is associated with a (1.345) times greater likelihood of the dependent variable or the outcome of interest occurring. Also note that the 95% C.I. of the odds ratio is from (1.254, 1.443).
A logistic regression analysis to investigate [Research Question] was conducted. The predictor variable, [Name of IV] was tested a priori to verify there was no violation of the assumption of the linearity of the logit. The predictor variable, [Name of IV], in the logistic regression analysis was found to contribute to the model. The unstandardized Beta weight for the Constant; $B = [wwww]$, $SE = [xxxx]$, $Wald = [yyyy]$, $p < .001$. The unstandardized Beta weight for the predictor variable: $B = [1www]$, $SE = [1xxx]$, $Wald = [1yyy]$, $p < .001$. The estimated odds ratio favored an [increase/decrease] of nearly [n%] $[Exp (B) = [zzzz]$, 95% CI (aaaa, bbbb)] for [Name of DV] every one unit increase of [Name of IV].
A logistic regression analysis to investigate if there is a relationship between Age and Retirement Planning was conducted. The predictor variable, Age, was tested a priori to verify there was no violation of the assumption of the linearity of the logit. The predictor variable, Age, in the logistic regression analysis was found to contribute to the model. The unstandardized Beta weight for the Constant; $B = (-2.503)$, $SE = 0.534$, $Wald = 21.948$, $p < .001$. The unstandardized Beta weight for the predictor variable: $B = 0.296$, $SE = 0.036$, $Wald = 67.892$, $p < .001$. The estimated odds ratio favored an increase of nearly 35% [$Exp (B) = 1.345$, 95% CI (1.254, 1.443)] for Retirement Planning every one unit increase of Age.
Summary

- As stated by Hsieh, Bloch, and Larsen (1998), in a simple logistic regression model, we relate a covariate $X_1$ to the binary response variable $Y$ in a model $\log \left( \frac{P}{1-P} \right) = \beta_0 + \beta_1 X_1$ where $P = \text{prob} (Y = 1)$. We are interested in testing the null hypothesis $H_0: \beta_1 = 0$ against the alternative $H_1: \beta_1 = \beta^*$, where $\beta^* \neq 0$, that the covariate is related to the binary response variable. The slope coefficient $\beta_1$ is the change in log odds for an increase of one unit in $X_1$.

Questions???
More questions? Stay Informed!

Current Walden students, e-mail our tutoring team at:  
AS Ctutoring@Waldenu.edu

Subscribe to our Facebook & Twitter channels to keep up-to-date on new information, ask questions and share your knowledge.

Subscribe to our YouTube channel (free!) to have instant access to new tutorials and webinar archives.

Visit our website for tutorials, event schedules, tutoring services, courses and workshops.