OR for multiple increments

Suppose, for neonatal survival, the adjusted

$OR_{10\text{ gms birthweight}} = 1.04$, and you want

$OR_{100\text{ gms birthweight}} = ?$.

First, calculate the coefficient for 100 gms and then calculate the OR for this new coefficient.

If $OR = e^b = 1.04$, the $b = \ln(1.04) = 0.0392$. For 100 gms, we need $10b = 0.392$ for the new coefficient.

Thus $OR_{100\text{ gms birthweight}} = e^{0.392} = 1.48$

Module 35: Logistic Regression Procedures

This modules describes the general logistic regression procedures. Several examples of the use of these procedures are included in a subsequent module.

Multiple Regression Background

For *Linear Multiple Regression*, we used the regression equation, or model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots ,$$

where the dependent variable Y was continuous and had a normal distribution.

For this situation, we were interested in how much of the variation of the dependent variable was explained by its relationship to the independent variables included in the equation.

Logistic Regression Situation

What can we do when the *dependent variable* has only two values, y = 1 indicating the person died and y = 0 indicating the person did not die.

For this situation, the *independent variables* could be exactly the same as those for a linear multiple regression model having a dependent variable that is continuous.

The logistic regression technique is ideally suited for this situation.

Logistic Regression Data Table

For the *independent variables*, the data table for a logistic regression is the same as a data table for the multiple regression situation.

The *dependent variable* has only two possible values, namely, $y = 1$ and $y = 0$, so that the entry for each person for this variable would either be 1 or 0.

Logistic Regression Data Table

	Death	Drug	Age	Weight	Sex
Person	y	x_1	x_2	x_3	x_4
1	1	1	47	160	1
2	0	0	53	140	0
3	1	0	61	130	0
⋮	⋮	⋮	⋮	⋮	⋮
n	1	1	52	180	1

Logistic Regression Equation

The logistic regression equation is:

$$\ln\{P/(1-P)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \cdots,$$

Where ln is the logarithm base *e*, often called the natural logarithm, with $e = 2.7183\cdots$, and

P = proportion in the population with $y = 1$,

e.g., proportion that died.

More on the Logistic Regression Equation

$$\ln\{P/(1-P)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots,$$

Note that

$$P/(1-P) = \text{Odds that } y = 1$$

So that the logistic regression equation focuses on ln(Odds).

Also, note that the independent variables can be the same as those for linear multiple regression.

The logistic regression equation can also be expressed, by exponentiation, as

$$\textbf{Odds} = \mathbf{P/(1-P)} = \mathbf{e}^{\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\ldots}$$

Adjusted Odds Ratios

Logistic regression models provide a way to obtain ORs that are adjusted for the terms in the model. For example, if we have a term in the model for sex, with

$$x = 1 \text{ for males}$$
$$= 0 \text{ for females}$$

and the "dependent" variable is

$$y = 1 \text{ if death}$$
$$= 0 \text{ if not a death}$$

We can obtain $OR_{m/f}$(death), adjusted for all the other terms that are in a logistic regression model.

Adjusted Odds for Males

The general equation for Odds is:

$$\mathbf{Odds} = \mathbf{P/(1-P)} = \mathbf{e}^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots}$$

For males, x = 1, so

$$\mathbf{O_m} = \mathbf{P_m/(1-P_m)} = \mathbf{e}^{\beta_0 + \beta_1 (X_1 = 1) + \beta_2 X_2 + \beta_3 X_3 + \dots}$$

$$= \mathbf{e}^{\beta_0 + \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \dots}$$

Adjusted Odds for Females

The general equation for Odds is:

$$\textbf{Odds} = \mathbf{P/(1-P)} = \mathbf{e}^{\beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\ldots}$$

For females, x = 0, so

$$P_f/(1-P_f) = \mathbf{e}^{\beta_0+\beta_1(X_1=0)+\beta_2 X_2+\beta_3 X_3+\ldots}$$

$$= \mathbf{e}^{\beta_0+0+\beta_2 X_2+\beta_3 X_3+\ldots}$$

Adjusted $OR_{m/f}$

$$OR_{m/f} = \frac{P_m/(1-P_m)}{P_f/(1-P_f)},$$

$$= \frac{e^{\beta_0+\beta_1+\beta_2 X_2+\beta_3 X_3+\ldots}}{e^{\beta_0+0+\beta_2 X_2+\beta_3 X_3+\ldots}}$$

$$= e^{\beta_1}$$

$$\widehat{OR}_{m/f} = e^{\beta_1}$$

Estimates of βs and Tests of Hypotheses

We have used the expression for the logistic regression equation for the population. The are computational procedures to use data from a sample to get estimates of the population parameters shown in the equation. The sample equation is:

$$\ln\{p/(1-p)\} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots,$$

There are also procedures for testing hypotheses about the model, starting with a test of the global hypothesis as per linear multiple regression and then proceeding on to the tests for individual parameters.

Confidence Intervals for β_i

The logistic regression computational programs provide estimates, b_i, and standard errors (SE) of the β_i. These can be used for confidence intervals for both the β_i and their associated ORs.

For the coefficients:

$$C\left[b_i - 1.96(SE) \le \beta_i \le b_i + 1.96(SE)\right] = 0.95$$

For the OR:

$$C\left[e^{b_i - 1.96SE} \le OR \le e^{b_i + 1.96SE}\right] = \mathbf{0.95}$$

The OR Hypothesis

The hypothesis of interest for ORs is:

H_0: OR = 1, vs

H_1: OR $\neq$ 1

Interpreting Adjusted ORs

If the adjusted $OR_{m/f} = 1.2$, then the odds for males being alive are 20% higher than the odds for females, adjusted for the other terms in the model.

If $OR_{m/f} = 0.75$, then the odds for males being alive are 75% of those for females, adjusted for the other terms in the model.

For a continuous independent variable such as birth weight, if $OR_{10\ gms} = 1.08$, then the odds ratio goes up 0.08 for every 10 gms increment in birthweight.