Newsom

Computing the Odds Ratio from Cell Frequencies

The crosstabs for the political party – gubernatorial candidate example is reprinted below.

party * candidat governor candidate Crosstabulation

			candidat governor candidate		Total
			.00 Dudley	1.00 Kitzhaber	
party	.00 Democrat	Count	15	30	45
		% within party	33.3%	66.7%	100.0%
		% within candidat governor candidate	27.3%	66.7%	45.0%
	1.00 Republican	Count	40	15	55
		% within party	72.7%	27.3%	100.0%
		% within candidat governor candidate	72.7%	33.3%	55.0%
Total		Count	55	45	100
		% within party	55.0%	45.0%	100.0%
		% within candidat governor candidate	100.0%	100.0%	100.0%

To compute the odds ratio, we need the following formula.

$$OR = \hat{\psi} = \frac{success\,vs.\ failure\ when\ X = 1}{success\,vs.\ failure\ when\ X = 0} = \frac{X = 1\,when\,Y = 1/\,X = 1\,when\,Y = 0}{X = 0\,when\,Y = 1/\,X = 0\,when\,Y = 0}$$

We will call support for Kitzhaber "success" in this example, because this candidate is coded Y = 1. Political party is the X variable, and we can use the frequencies of the four cells to compute the odds ratio.

$$\begin{aligned} OR &= \frac{X = 1\,when\,Y = 1/\,X = 1\,when\,Y = 0}{X = 0\,when\,Y = 1/\,X = 0\,when\,Y = 0} \\ &= \frac{15/40}{30/15} \\ &= \frac{.375}{2} \\ &= .188 \end{aligned}$$

This value matches the odds ratio, or the exp(B) value, obtained from SPSS, and suggests that the odds that Republicans are over five times less likely (1/.188 = 5.319) to support the Kitzhaber than Democrats.

Computing Predicted Probabilities from the Regression Equation

To compute the predicted probability for any value of the outcome, the exponent function is needed to transform the regression equation from log values back to probabilities. The general formula is:

$$\hat{p} = \frac{1}{1 + e^{-(B_1 X_i + B_0)}}$$

Where the predicted probability for Y for any particular value of X can be computed if the values of X, B_1, and B_0 are inserted into the formula.

To start with something simple, one can compute the proportion of successes or Y = 1. This is the same as the average for the full sample. Using the intercept (i.e., constant) from the logistic output for the gubernatorial simple logistic example obtained from the Block 0 step in which there are no predictors in the model, I compute the probability of Y = 1, which is the proportion overall who supported Kitzhaber. The constant value was -.201, so the probability of support for Kitzhaber in the full sample was:

$$\begin{aligned}\hat{p} &= \frac{1}{1 + e^{-(B_0)}} \\ &= \frac{1}{1 + e^{-(-.201)}} \\ &= \frac{1}{1 + e^{.201}} \\ &= \frac{1}{1 + 1.223} \\ &= .45\end{aligned}$$

Which is equal to the overall proportion supporting the Kitzhaber obtained from the crosstabs procedure.

Similarly, we can use the full equation to obtain the predicted probability that Y = 1 for any particular value of X. Using the second logistic regression example with years experience predicting success in the widget business, I calculated the predicted probability of success with 10 years of experience.

$$\begin{aligned}\hat{p} &= \frac{1}{1 + e^{-(B_1 X_i + B_0)}} \\ &= \frac{1}{1 + e^{-(.190*10 + (-1.968))}} \\ &= \frac{1}{1 + e^{(-1.90 + 1.968)}} \\ &= \frac{1}{1 + e^{-.068}} \\ &= \frac{1}{1 + .934} \\ &= .517\end{aligned}$$

To check, we can inspect the graph obtained from the logistic output. When years experience is equal to 10, the predicted probability is approximately this value.