Print

Calculating Canonical Links for Binary Response and Count Data

To identify the canonical link for binary data, start with the mass function of a Bernoulli (binary) random variable and rearrange to the following format (shown previously):

$$f(y \mid \theta) = h(y) \cdot c(\theta) \cdot e^{\sum t_i(y_i)(W_i(\theta))}$$

$$f(y \mid p) = p^y \cdot (1-p)^{(1-y)} = p^y \cdot (1-p)^1 \cdot (1-p)^{-y}$$

$$= \frac{p^y \cdot (1-p)}{(1-p)^y} = (1-p) \cdot I_y \cdot \left(\frac{p}{1-p}\right)^y = (1-p) \cdot I_y \cdot e^{\log\left(\frac{p}{1-p}\right)^y}$$

$$= (1-p) \cdot I_y \cdot e^{y \cdot \log\left(\frac{p}{1-p}\right)} \qquad \text{where } I_y = \begin{cases} 1 & \text{for } y = 0,1 \\ 0 & \text{otherwise} \end{cases}$$

In this format you can see that the canonical link is $\left(\frac{p}{1-p}\right)$

In PROC GENMOD, the location parameter is *p*, and the scale parameter is 1.

The notation ***log*** refers to the natural logarithm, which is the logarithm to the base e. The log link function is often applied to count data that have nonnegative integer values. The log transformation removes the lower bound and creates a linear model.

The probability mass function for Poisson distribution is given by the following:

$$f(\mathrm{y} \mid \lambda) = \frac{e^{-\lambda} \lambda^y}{y!} \qquad y = 0, 1, 2, \cdots$$

In this expression, *y* is a nonnegative integer value and *λ* is the expected value of *Y*. It can be shown that Var(*y*) = *λ* and the scale parameter is 1.

To identify the canonical link, rearrange it to the following format: $f(y \mid \lambda) = h(y) \cdot c(\lambda) \cdot e^{t(y)(W(\lambda))}$

$$f(y \mid \lambda) = \frac{e^{-\lambda} \lambda^y}{y!} = \frac{e^{-\lambda} \cdot e^{\log(\lambda^y)}}{y!} = e^{-\lambda} \cdot \frac{1}{y!} \cdot e^{y(\log(\lambda))}$$

.

The last term indicates that the log is the canonical link. In PROC GENMOD, the location parameter is λ and the scale parameter is 1.

Close