Lecture 38

LOGISTIC REGRESSION - AN EXAMPLE

This lecture discusses the Minitab output and interpretation of a model of survivorship for adults in the Donner party. The data are available at:

`http://www.stat.psu.edu/online/development/stat504/06_logreg/graphics/donner.txt`

The first column is age, the second is an indicator variable for sex (1 = male, 0 = female) and the third is the response (1 = survived, 0 = died).

The Donner party were trapped in the Sierra Nevada mountains in a sudden early snowstorm on their way to California. They were not well equipped to handle the situation - they lacked food and had only rough shelter. The trees were too large to be chopped down easily so they also lacked timber. Many perished and the party is also famous because of rumors of cannibalism.

The model was found using the Binary logistic regression option under Stat > Regression.

Logistic Regression Example

A full model with age, sex, and age by sex interaction gives:

$$\text{logit}(\pi) = 7.24 - 0.194 \text{ age} - 6.928 \text{ male} + 0.1616 \text{ male} \times \text{ age}$$

which gives the odds of surviving as

$$\frac{\pi}{1-\pi} = e^{7.24} e^{-0.194 \text{ age}} e^{-6.928 \text{ male}} e^{0.1616 \text{ male} \times \text{age}}$$

$$= 1403.02(.82)^{\text{age}} (0.00098)^{\text{male}} (1.175)^{\text{male} \times \text{age}}$$

The interpretation when there are interaction terms in the model is less straightforward than when the model is additive and has to be made separately for each sex and age. For women, male is zero in the above model. So each year increase in age multiplies the odds of survival by 0.82. This means that, for a woman, every year increase in age is associated with an 18% decrease in her odds of survival.

For a man, each year increase in age multiplies the odds of survival by $(\mathbf{0.82})(\mathbf{1.175}) = \mathbf{0.9635}$. Thus, for a man, each year increase in age is associated with a 3.65% decrease in his odds of survival.

To compare the odds a man survives to the odds a woman survives requires knowing something about the age. The number 0.00098 is for when age is zero which extrapolates beyond the scope of the data. The individuals in the study range from age 15 on up to ages in the 60's. Given an age, the odds ratio for men to women is:

$$\frac{\text{odds a man of a certain age survives}}{\text{odds a woman of the same age survives}} = (\mathbf{0.00098})(\mathbf{1.175})^{\text{age}}$$

This odds ratio for 30 years old is 0.12, meaning that the odds a 30 year old man survived is just 12% of that of a 30 year old woman. For 50 years old, the odds ratio is 3.11, meaning that the odds a 50 year old man survived is 3 times that a 50 year old woman survived.

REFERENCES AND READINGS

Exercises for Lecture 38

1. –   2. –