Logistic Regression (with a single explanatory variable)

STAT 340: Applied Regression

Exercise 1: The linear probability model (Fox, 3rd Ed., Section 14.1)

A plebiscite is defined by Merriam Webster as "a vote by which the people of an entire country or district express an opinion for or against a proposal especially on a choice of government or ruler". In September 1988, the Chilean people participated in a plebscite to decide whether to support the military government, established 15 years earlier in a coup. A yes vote supported 8 more years of military rule, while a no vote supported a process to return to civilian rule. A national survey of 2700 randomly selected Chilean voters was conducted six months before the plebiscite; 868 planned to vote yes, 889 planned to vote no, and the remaining were undecided or abstained.

In your textbook, the author looks at the relationship between support for the status quo, a scale formed from a variety of questions on political, social, and economic policies (high scores mean support for the military regime policies), and voting intention. The range for status quo is between -2 and 2.

Consider the least-squares line fit to the Chilean plebiscite data (Section 14.1, J. Fox):

$$\hat{\pi}_{yes} = 0.492 + 0.394 \times \text{Status quo}$$

where $\hat{\pi}_{yes}$ is the estimated probability of a yes vote.

Would you recommend this model? If so, why? If not, what are some potential problems?

Exercise 2: Daycare Centers and Respiratory Health (Nafstad et al. 1999)

Interpret the odds ratio in the following abstract.

Objective. To estimate the effects of the type of day care on respiratory health in preschool children.

Methods. A population-based cross-sectional study of Oslo children born in 1992 was conducted at the end of 1996. A self-administered questionnaire inquired about day care arrangements, environmental conditions, and family characteristics (n = 3853; response rate, 79%).

Results. In a logistic regression controlling for confounding, children in day care centers had more often nightly cough (adjusted odds ratio, 1.89; 95% confidence interval 1.34-2.67), and blocked or runny nose without common cold (1.55; 1.07-1.61) during the past 12 months compared with children in home care.

Exercise 3: Soccer goals on target (Roskes et al. 2011)

Data comes from an article in Psychological Science (Roskes et al. 2011). The authors report on the success rate of penalty kicks that were on-target, so that either the keeper saved the shot or the shot scored, for FIFA World Cup shootouts between 1982 and 2010. They found that 18 out of 20 shots were scored when the goalkeeper's team was behind, 71 out of 90 shots were scored when the game was tied, and 55 out of 75 shots were scored with the goalkeeper's team ahead.

Calculate the odds of a successful penalty kick for games in which the goalkeeper's team was behind, tied, or ahead. Then, construct empirical odds ratios for successful penalty kicks for (i) behind versus tied, and (ii) tied versus ahead.

References

J. Fox. 2016. *Applied Regression Analysis and Generalized Linear Models*, 3rd Edition. Sage.

P. Roback and J. Legler. 2020. *Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R.*