

LOGISTIC MODEL DATA

Contents

1	Data	1
1.1	Home Situation	1
1.2	Academic Indicators	1
1.3	Testing Effects	1
1.4	Retention	1
2	Area Models	2

1 Data

The data provided consisted of 111 observations which were each comprised of 23 variables. The variables `ATTEND`, `BEHAVE2`, `LETREC2C`, `LETREC2L`, and `NOREC2` were discarded because they contained information which would not be available at the time the parent was attempting to determine whether a prospective student is ready for kindergarten.

To reduce the effects of sparseness in the modeling process, the remaining variables were grouped into three areas: home situation, academic indicators, and testing effects. A short description of each of these groups follows.

1.1 Home Situation

The variable `BRTHORDR` is a discrete variable giving the birth order of the child. `TOTCHILD` is another discrete variable indicating the total number of children in the household.

`GENDER` is a categorical variable which was coded as 0 for *females* and 1 for *males*. `RACE` is another categorical variable which was coded as 1 for *black*, 2 for *white*, 3 for *hispanic*, 4 for *Pacific islander*, 5 for *Asian*, 6 for *Filipino*, and 7 for *other*.

The variable `BILING` is a categorical variable representing whether the child is eligible for bilingual instruction (0=*no*, 1=*yes*), and was viewed as representing whether more than one language was commonly used in the home.

The final variable in this group is `AGE1290` which was computed from `BDAYMO` and `BDAYR`. It represents the age of the child, in months, in December of 1990.

1.2 Academic Indicators

The variables `LETREC1C` and `LETREC1L` are discrete variables representing the number of letters (capital and lower case respectively) recognized by the child. Similarly, `NOREC1` is a discrete variable representing the number of numbers recognized.

`ROUND2` is a categorical variable (0=*no* and 1=*yes*) which indicates whether the student will be

in kindergarten for a second time—*i.e.* the student was previously retained.

Finally, `BEHAVE1` is an ordinal variable (1=*good*, 2=?, 3=?, and 4=*bad*) representing the teacher's perception of the students behavior at entry. While this value was not assessed prior to the child's entry, it could easily be assessed earlier.

1.3 Testing Effects

Information on both teacher and school were recorded to control for these effects. The variable `TEACHER` is a categorical variable which indicates which of the four teachers the child was assigned to. `SCHOOL` is a categorical variable used to indicate which school the child attended. Due to the method of sampling—only one teacher was used at Ramona—teacher three was confounded with school two.

1.4 Retention

The most important variable in the group was `RETAINED`. This dichotomous categorical variable (0=*no* and 1=*yes*) indicated whether the teacher felt the student should be retained. Actual retention could not be used as the outcome because of legal issues.

2 Area Models

As noted above, the three areas were modeled separately to reduce the effects of sparseness in the logistic regression model.¹ Variables which survived the selection process were then tested for interactions.

References

- [1] A. Agresti. *Categorical Data Analysis*. John Wiley & Sons, New York, 1990.
- [2] J. L. Fleiss. *Statistical Methods for Rates and Proportions*. John Wiley & Sons, New York, 1973.
- [3] David W. Hosmer, Jr. and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, New York, 1989.

¹Logistic regression is a technique which allows for the fitting of multidimensional contingency tables. While dichotomous dependent variables are the norm, logistic regression may also be used to fit ordinal outcomes. Another advantage of the use of logistic regression is its ability to incorporate interval-ratio (discrete or continuous) covariates into otherwise purely categorical models.

Fleiss[2] discusses the relationship of logistic models and odds-ratios. For information and logistic regression modeling and validation see Hosmer and Lemeshow[3]. Agresti[1] gives a more technical, but thorough description of the various forms of logistic regression.