SAN JOSÉ STATE UNIVERSITY **College of Social Work**
S. W. 242
Spring 2009
Edward Cohen

Week 13 Class Notes

- Go over test questions
- Introduction to Logistic Regression
- Final Paper requirements
- Affinity group work

I. Logistic Regression

A. Definition and Types

1. *Binomial (or binary) logistic regression* is a form of multivariate regression which is used *when the dependent variable is a dichotomous categorical variable and the independent variables are of any type*. The DV (e.g. cured/not cured; happy/not happy; used service/didn't use service) is coded as a dummy variable (1 = Yes, 0 = No).

2. Although the outcome or dependent variable is dichotomous, logistic regression is similar to multivariate linear regression in that the purpose is to analyze the effects of the predictor variable(s) on the dependent variable, controlling for all other IVs. In this case, it's *the effect of the predictor(s) on whether the outcome occurs or not.*

3. Extensions of the logistic model include *multinomial logistic regression*, used when there are more than two nominal categories in the DV (such as neglected, sexually abused, or physically abused). *Ordinal logistic regression* is used when the DV is ordinal (such as Very happy, moderately happy, moderately unhappy, and very unhappy). In this class we will only discuss the binomial logistic regression model.

B. The use of logistic regression started in biostatistics and epidemiology. Logistic regression is now used in many social and hard science research fields. Its popularity exploded after the publication of: David W. Hosmer & Stanley Lemeshow (1990). *Applied logistic regression*. New York: Wiley.

C. Typical research questions for binomial logistic regression:

1. "What is the likelihood of having a suicide attempt, controlling for number of previous suicide gestures, severity of depression, ethnicity, gender?"

2. "Does the successful completion of a substance abuse treatment program depend on higher levels of motivation? What role do ethnicity, gender and type of drug play?"

3. "Do support groups for released felons prevent being re-arrested, controlling for number of previous arrests, seriousness of previous crime, and ethnicity?"

4. "Can reunification home after a child abuse report be predicted by type of abuse allegation, satisfaction with services, ethnicity, and previous child abuse reports?"

D. The logic of logistic regression

1. Since the DV is not continuous, the assumption of linearity (i.e. that the IV is linearly related to the DV) does not apply as it does for bivariate and multivariate linear regression.

2. Logistic regression estimates *the odds of a certain event occurring* (the DV event), or how likely it is (the odds) that the observed values of the DV may be predicted from the observed values of the IVs. (See below for definitions of odds and odds ratios.)

3. In research articles you might encounter the term "*maximum likelihood*." This is the statistical procedure used in logistic regression (it doesn't use least squares regression). Maximum likelihood picks the values of the model parameters (e.g. the beta coefficients) that make the observed data "more likely" than they would be with other parameters. The statistical procedure tries several "iterations" until reaching the best fitting parameter estimates. Making interpretations of the basic statistical output does *not* require knowledge of how maximum likelihood works.

Example of Logistic Regression: "Partnering with Parents in Child Welfare: A Promising Practice for Reunification in Public Child Welfare"

This mixed methods study looked at reunification outcomes for parents assigned a Parent Partner. Parent Partners are former child welfare clients (for the most part Latino or African American) who successfully navigated the system, and were later hired by the County to provide mentoring, informal support, and advocacy to new birth parents facing court-ordered placement of their children. The study site is Contra Costa County. The research design was a pre-experimental posttest only, with non-equivalent comparison group. From a final sample of 113, 56 clients who benefited from parent partner services (beyond one meeting for information and referral) were matched with 57 parents of closed cases from 2007. Parents were matched on ethnicity, type of abuse, gender, and time in the child welfare system. It was hypothesized that having a Parent Partner would result in higher rates of reunification than for parents in the comparison group. Variables available for analysis were:

- Group (Parent Partner Group = 1; Comparison Group = 0)
- Dummy variable for Latino (1 = Latino 0 = Not Latino) with White as the reference category
- Dummy variable African American (African American = 1; Not African Amerian = 0), with White as the reference category
- Child's age at removal, in years
- Child's gender
- Reunification within 18 months (1 = yes; 0 = no)

1. Identify the independent variables and level of measurement.

- Group – categorical

- Latino – categorical
- African American – categorical
- Age at removal – continuous
- Child's gender – categorical

2. Identify the dependent variable and level of measurement.

- Reunification -- categorical

3. What is the null hypothesis for the overall model fit?

The overall model does not predict reunification (or, the IVs as a group are not related to reunification).

For the Independent variables?

The separate IVs are not related to the likelihood of reunification.

4. What is the alternative hypothesis for the overall model fit?

The overall model predicts the likelihood of reunification.

For the independent variables?

Those having a Parent Partner are more likely to be reunified within 18 months, controlling for age, gender and type of abuse. Ethnicity may also be related to reunification.

5. Why is logistic regression suitable for these data? Identify the alpha level.

Categorical (dichotomous) DV with any type of IVs. Alpha $\leq$ .05

6. Review SPSS output.

SPSS procedure: **Analyze➔Regression➔Binary logistic; enter variables; click Options and check Hosmer-Lemeshow goodness-of-fit**

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.839	8	.554

Just like linear regression, we need to have some way to know if the model is sufficient; if it "fits the data." There are several options to test this. This one, the **Hosmer and Lemeshow Test**, tests the Null that the *observed* probability of reunification for each case is no different than that *predicted* by the model, dividing the cases into ranked groups and comparing their probabilities using Chi Square. If the observed and predicted probabilities are nearly the same, this indicates that the model does a good job at explaining the variance in the dependent variable (reunification). Here, the *p* value of .554 is far greater than our alpha $\leq$ .05. *We cannot reject the Null, which is good news for our regression model.* The model (the group if independent variables) fits the data, and we can move on to see which variables are likely to influence reunification.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Age_at_removal	-.005	.046	.011	1	.917	.995
	Gender(1)	-.597	.399	2.242	1	.134	.550
	Black	-.568	.489	1.349	1	.246	.567
	Hispanic	-.329	.512	.412	1	.521	.720
	group	.924	.400	5.321	1	.021	2.518
	Constant	-.012	.413	.001	1	.977	.988

This table shows the independent variable coefficients and related statistics. This requires some explanation.

B is the regression coefficient in logistic regression, but it is not interpreted the same as in linear regression. While still showing the impact of a one unit change in the IV on the DV, controlling for all other variables, this B is a "log odds," e.g. having a Parent Partner increases the log odds of reunification by .924 (hence the name "logistic" regression). Not a very helpful statistic!

The last column, **Exp(B)** is the "exponentiated B" (taking the B value and calculating the inverse natural log[1] of B) which is, conveniently, an odds ratio. **Odds** means the probability of an event occurring, divided by the probability of the event not occurring. Example: let's say the probability of being randomly selected in class to answer a question is 1 out of 25, or .04. The probability of <u>not</u> being selected is (1 - .04) = .96 (the same as 24/25). The odds of being selected are .04/.96 = .04167, rounded to .04.

Now let's say you are also in another class with 20 students. Your odds of being randomly selected there are 1/20 divided by 19/20, = .05/.95 = .0526, rounded to .05. An **Odds Ratio** is the odds of an event occurring in one group, divided by the odds of the event occurring in another group. So your odds ratio of being selected in the first class vs. the other is .04/.05 = 0.8, which means the odds of being selected in the first class are 80% that of the second class. If you reversed the numerator and denominator to .05/.04 = 1.25, you would have a 25% greater odds of being selected in the second class. The odds ratio in binomial logistic regression and with a dichotomous (categorical) independent variable, is the *odds of an event occurring if the independent variable is coded 1, divided by the odds of an event occurring if the independent variable is coded 0.*

Now let's look at the odds ratio of reunifying vs. not reunifying, as affected by having a Parent Partner. The odds ratio is the odds of reunification for Group = 1 (having a Parent Partner), divided by the odds of reunification for Group = 0 (not having a Parent Partner). In the output, the Exp(B) = 2.518. Interpreting this, we can say that having a Parent Partner increases the

[1] The "natural logarithm" of a number x is the power to which *e* would have to be raised to equal x. For example, the natural log of *e* itself is 1, since $e^1 = 1$. The natural log of 1 would be 0, since $e^0 = 1$.

likelihood or odds of reunification within 18 months by more than 2.5. Alternatively, we can say that those with a Parent Partner are more than twice as likely as those in the comparison group to reunify within 18 months, controlling for age at removal, gender, and ethnicity.

An Exp(B) over 1.0 signifies that the independent variable *increases the odds* of the dependent variable occurring (whatever was coded 1 in the DV). An Exp(B) under 1.0 signifies that the independent variable *decreases the odds* of the dependent variable occurring. Another clue about the direction of effect can be seen in B — a negative B coefficient will result in an Exp(B) less than 1.0, and a positive B coefficient will result in an Exp(B) greater than 1.0.

The statistical significance of each B is tested by the **Wald Chi-Square**—testing the null that the B coefficient = 0 (the alternate hypothesis is that it does not = 0). *p* values lower than alpha are significant, leading to rejection of the null (just like in linear regression).

For independent variables that are continuous, the interpretation of the Exp(B) is almost the same. In this example, the Exp(B) for age = .995. A one unit increase in age results in a decreased odds of reunification of .995. This is a trivial amount of decrease. (An Exp(B) coefficient of 1 means the odds do not change at all.) As it turns out, the coefficient for Age is not statistically significant anyway.

7. Results—decision to accept/reject null

The Hosmer Lemeshow test of model fit shows that the overall model is predictive of reunification , X^2 (8) = 6.84, *p* = .554, and the null that the model is predictive cannot be rejected.

We can reject the null that the B coefficient for having a Parent Partner is equal to zero—it is statistically significant and predictive of reunification (Exp(B) = 2.52, *p* = .021). Those parents assigned to a Parent Partner are more than twice as likely to reunify within 18 months, controlling for age at removal, ethnicity and gender. Reunification was not influenced by age at removal, ethnicity, and gender in this sample.

8. Discussion of Results

This preliminary study in the effectiveness of a parent mentoring and advocacy program supports the hypothesis that having a Parent Partner increases the chances of reunification for birth parents, controlling for ethnicity, age at removal, and gender. Providing informal peer support for birth parents facing removal of a child can help them navigate the child welfare system and eventually reunify with their children. There are important limitations to this study. A non-probability sampling procedure was used to select parents in both groups, adding the potential for selection bias in sampling. Early families in the program who were eligible for the study may have different characteristics than most others currently receiving services. The pre-experimental design and comparison group matching may not have taken into account other important variables, such as having a mental illness or developmental disability, or a history of previous abuse reports. Further studies with larger samples and a more rigorous research deisgn would be required to address these limitations. Despite these limitations, parent mentoring holds promise

as a way to successfully intervene with Latino and African American families who need help navigating through a highly adversarial and stressful situation.