WWS 509 Generalized Linear Models: Precept 4
Section 3.5

Kristin E. Bietsch

Office of Population Research, Princeton University

October 2012

1 Introducing the Data

Today we will be looking at the same data as last week, women giving birth after HIV testing. For today, I have created group data, based on age groups (15-19, 20-24, 25-29, 30-34, 35-39, 40-44) and HIV status.

2 Deviance

Deviance is the goodness of fit likelihood ratio chi-square statistic

- With a degree of freedom determined from the number of parameters added to the model
- To see if a model fits, you want to look at the deviance and check the χ^2 distribution.

Regardless if the group or individual data, the estimated effects, the standard errors, and likelihood ratio test based on differences between deviances will be the same (but the deviances will change).

3 Categorical Age and Binary HIV status

There are 5 potential models from this data, what are they?

Model	Notation	logit(π_{ij})	Deviance	D.F.
Null	ϕ	η	366	11
Age	A	$\eta + \alpha_i$	38	6
HIV	H	$\eta + \beta_j$	353	10
Additive	A+H	$\eta + \alpha_i + \beta_j$	4	5
Saturated	AH	$\eta + \alpha_i + \beta_j + (\alpha\beta)_{ij}$	0	0

Using the above table:

1. How would I test the gross effect of age?

   - Compare the model with age to the null. There is a difference in deviance of 328 costing 5 degrees of freedom. This is obviously significant!

   (a) Test if the model including only age fits the data.

   - The model has a deviance of 38 on 6 degrees of freedom. This is significant, which means our model does not yet fit the data.

2. How would I test the gross effect of HIV?

   - Compare the model with HIV to the null. There is a difference in deviance of 13 costing 1 degrees of freedom. This is obviously significant!

   (a) Test if the model including only HIV fits the data.

   - The model has a deviance of 353 on 10 degrees of freedom. This is significant, which means our model does not yet fit the data.

3. Which variable, HIV or age, explains more (only looking at the single variable models)?

   - It looks like age explains more, it reduces the deviance so much!

   (a) Can you test this?

   - No! The models are not nested, so we cannot test the significance between them.

Now look at the additive model in the table:

1. What can this model tell you?

   - Here, we can look at the addative effects of HIV and age. We can look at the effect of one, holding the other constant. For example, we can look at the difference age makes for women of the same age category. Or we can look at the difference HIV makes in childbearing for women of the same age group. We are assuming that HIV has the same effect at each age.

   (a) What is the effect of adding age to the model containing HIV status?

   - We reduce the deviance from 353 to 4, a reduction of 349, costing 5 degrees of freedom. This is significant, meaning age improves the model.

   (b) What is the effect of adding HIV to the model containing age?

- We reduce the deviance from 38 to 4, a reduction of 34, costing 1 degrees of freedom. This is significant, meaning controlling for HIV status improves the model.

2. Does this model fit the data?
   - Yes! We have a deviance of 4 with 5 degree of freedom. This has an insignificant chi-square!

Now finally, look at the saturated model in the table:

1. How much deviance is in this model?
   - Zero, it is saturated
2. Is this model useful for interpretation?
   - Not really, it basically just gives each groups mean.

4 Linear Age and Binary HIV status

Now age as a categorical variable uses up many of our degrees of freedom, so lets see if we can model things a bit differently. Refer to the output in Appendix B. Fill in the deviance table for all the models in Appendix B (and don't forget to include the null model from Appendix A).

Model	Notation	logit(π_{ij})	Deviance	D.F.
Null	ϕ	η	366	11
Age (linear)	X	$\alpha + \beta X_i$	362	10
Age (linear)+ HIV	$X + H$	$\alpha_j + \beta X_i$	347	9
Age (linear)+ HIV + Interaction	$X + H + XH$	$\alpha_j + \beta_j X_i$	345	8
Age (linear + square)	$X + X^2$		41	9
Age (linear + square)+ HIV	$X + X^2 + H$		6	8

1. What do you think about modeling age linearly?
   - It doesn't reduce the deviance as much as using grouped data (but it uses less degrees of freedom). We cannot do a formal test between the two models since they are not nested.

   (a) Is it an improvement over the null?
   - It reduces the deviance by 4 at the cost of 1 degree of freedom, which is significant at the 5% level. So age stays in.

   (b) Does this model fit the data?
   - No

I added HIV into the model:

1. Is this model an improvement over the previous model?

   - There is a reduction of 5 in the deviance at the cost of 1 degree of freedom. Thus, this is an improvement.

   (a) Does this model fit the data?

   - No, this model does not come close to fitting the data.

Lets now try an interaction term?

1. Is the interaction term significant (look at the output)?

   - No.

   (a) Does this model offer an improvement over the additive model?

   - No, there is a reduction of deviance of 2 at the cost of 1 degree of freedom. This is not statistically significant on the chi-square distribution.

OK, linear effect of age and childbearing, not a good idea. But this doesn't mean we have to return to our categorical variables. I add a quadratic term for age.

1. How did I create this term in Stata?

   - I centered the interaction term around age 27.5 to reduce collinearity with age.

2. What do you think about this model?

   - This is a fantastic model! We can compare it to the model of age as linear since they are nested. This is a reduction in deviance of 321 on only 1 degree of freedom!

   (a) Does this model offer an improvement over the model of age as a linear predictor?

   - We can compare it to the model of age as linear since they are nested. This is a reduction in deviance of 321 on only 1 degree of freedom!

For the last regression, I model age, age squared, and HIV status.

1. Does this model offer an improvement over the model of age and age squared?

   - Yes, we reduce the deviance by 35 with only 1 degree of freedom.

2. Does this model fit the data?

   - Yes, we only have a deviance of 6 with 8 degrees of freedom.