

Homework #4

Capture Recapture Methods

- The following data were obtained to estimate the number of births with spinal bifida. There were three sources of data. For example there are 13 people who are in Source 1, but not in Sources 2 and 3.

		Source 1*				x
		Yes		No		
		Source 2*		Source 2		
		Yes	No	Yes	No	
Source 3*	Yes	0	3	1	8	$N_{\text{obs}} = 36$
	No	8	13	3	0	

Using this data, please do the following:

- Read the data in to R (or SAS), in a similar way to what we did in class for the four data source example. (Hint: you should have seven rows and four columns in your data set.)
 - First fit a log linear model that assumes that all the data sources are independent of each other. What is your estimate for the total number of births with spinal bifida?
 - Now fit a model that assumes that there is dependence between sources 1 and 2 and sources 2 and 3. What is your estimate for the total number of births with spinal bifida based on this model?
 - IF** the model that you fit in part c is the most correct model, what does this tell you about source dependence in this setting? Does it exist? If so, what kind is there? (note that this may not be the best model, but the goal here is to just interpret the implications of this model)
- What is capture recapture used for? What are alternative methods to capture recapture?