BIOS6643 Longitudinal

L17 Ordinal logistic regression

EJC

Department of Biostatistics & Informatics

Ordinal regression

Case study

Logistic regression

Summary

Topics for these notes:

- Ordinal logistic regression

Associated reading: as of now, the course notes do not have more information than what is here.

Case Study

"Since 2001, 3 million soldiers have deployed to Southwest Asia (SWA), with exposure to inhalants that cause respiratory disease. Department of Defense uses standard occupational codes, termed Military Occupational Specialty (MOS), to classify military personnel by job/training. We characterized Marine MOS by estimated exposure to inhalational hazards. We developed an MOS-exposure matrix containing five major deployment inhalational hazards–sandstorms, burn pits, exhaust fumes, combat dust, occupational VDGF (vapor, dust, gas, fumes)–plus time worked outdoors. A 5 member expert panel of two physician deployment veterans and three occupational pulmonologists independently ranked 38 Marine MOS codes for estimated exposure intensity (3=high, 2=medium, 1=low) to each hazard." From Pepper et al., 2017.

The MOS occupational codes (or MOS_num) are numbered 1 through 38, for convenience, but they relate to specific job types. For example, 1=personnel and administration, 2=intelligence, 3=infantry, etc.

Our data follows this form, for a given inhalation hazard:

Rater	MOS1	MOS2	MOS3	MOS4	MOS5	...
1	1	1	3	2	1	
2	1	1	3	1	1	
3	1	2	3	2	1	
...						

The outcome is ordinal and given that there are only 3 levels (3 is high exposure, 2 is medium, 1 is low), we consider a model that is specialized for this type of outcome.

A GzLMM that can be used to fit our data has the form
$\lambda_{ijk} = log\left[\frac{P(Y_{ij} \leq k \mid b_i,\ b_j)}{1-P(Y_ij \leq k \mid b_i,\ b_j)}\right] = \alpha_k + b_i + b_j,$

where $i =$ MOS_num, $j=$ rater, and k is outcome level;
$\alpha_k,\ k = 1,\ ...,\ K-1$ are strictly increasing intercepts; bi and bj are random intercepts for MOS_num and rater, respectively.

In order to get estimates that are commensurate with increasing levels of the outcome, we can reverse the inequalities to obtain
$\lambda^c_{ijk} = log\left[\frac{P(Y_{ij} \geq k \mid b_i,\ b_j)}{1-P(Y_{ij} \geq k \mid b_i,\ b_j)}\right] = \alpha_k + b_i + b_j.$

This is the model we will fit for the application. We achieve this model using a 'descending' option, discussed shortly.

Some questions of interest for our data:

1. How do variances for raters compare with the variances over MOS types?
2. Are there any raters that significantly differ from the group average?
3. After adjusting for crossed random effects of MOS type and rater, what are the cumulative odds of low, medium, high exposure for a given inhalation hazard?
4. What is the probability of a particular job of having a high exposure to a given exposure type?

To answer these questions, we can fit the ordinal logistic regression model shown on the last slide that accounts for multiple measures per MOS type (called MOS_num below), which is the experimental unit here (instead of subjects).

Descriptive approach to obtaining probabilities and odds ratios

First, to get an understanding of the statistics we're dealing with, let's consider the data more descriptively.

In the data, we have 115 MOS's assigned as 'low exposure' job types (62.5%), 56 as 'medium' (30.4%) and 13 as 'high' (7.1%).

Without considering the correlation, the odds of a medium or high classification for a randomly selected MOS is (0.375)/(1-0.375) = 0.6; the odds of a high classification is 0.071/(1-0.071)=0.076.

When we fit the model, we account for the fact that rater's score every MOS; i.e., the random effects are crossed. (In a previous data set we talked about raters and subjects being crossed, e.g., 'Dancing with the Stars'.) This may impact the results.

Back to the ordinal logistic regression

SAS Code for one inhalation exposure source, burn pits:

```
proc glimmix data=all2 method=laplace;
 class mos_num rater;
 model burn_pits(desc) =
  / solution dist=multinomial link=cumlogit;
    random mos_num rater / solution;  run;
```

The 'desc' option is added so that the direction of estimates and outcome levels are consistent.

Note that in the model statement, there are no effects; thus we only have the intercepts. Generally for logistic regression there are K-1 intercepts for an outcome with K levels; here, K=3, so we'll get 2 intercepts. In other applications you might add covariates.

The Laplace method approximates the true likelihood, and hence considered ML estimation.

The GLIMMIX Procedure

Model Information

Data Set	WORK.ALL2
Response Variable	Burn Pits
Response Distribution	Multinomial (ordered)
Link Function	Cumulative Logit
Variance Function	Default
Estimation Technique	Maximum Likelihood
Likelihood Approximation	Laplace
Degrees of Freedom Method	Containment

Number of Observations Used 184

Response Profile

Ordered Value	Burn_Pits	Total Frequency
1	3	13
2	2	56
3	1	115

The GLIMMIX procedure is modeling the probabilities of levels of Burn_Pits having lower Ordered Values in the Response Profile table.

The intercept for Burn_Pits=2 means that the associated odds ratio will be for levels 2 or 3, relative to 1; the intercept for Burn_Pits=3 compares 3 versus 1 and 2.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error
MOS_num	2.9181	1.2889
rater	0.7259	0.6157

The variance estimates indicate that the variability of the exposure estimates among job types (MOS_num) is 4 times greater than for the raters, which is probably reassuring to the raters.

Solutions for Fixed Effects

Effect	Burn_Pits	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	3	-3.8512	0.6760	4	-5.70	0.0047
Intercept	2	-0.7868	0.5219	4	-1.51	0.2062

The odds of a rater ascribing a job type as having medium or high exposure (relative to low) is exp(-0.7868)=0.46; the odds of high versus medium or low is exp(-3.8512)=0.02. Even with the Total Frequency table above, we see that 3's (i.e., 'High' exposure) are more rare.

Solution for Random Effects

Effect	rater	MOS_num	Estimate	Std Err Pred	DF	t Value	Pr > |t|
MOS num		1	-0.5945	0.9462	142	-0.63	0.5308
MOS_num		2	0.1359	0.8699	142	0.16	0.8761
MOS_num		3	3.2523	1.0217	142	3.18	0.0018
…							
MOS num		73	0.09849	0.8591	142	0.11	0.9089
rater	Gottschall		-1.1538	0.5745	142	-2.01	0.0465
rater	Kreft		0.3367	0.4953	142	0.68	0.4977
rater	Meehan		0.4260	0.4993	142	0.85	0.3951
rater	Pepper		0.9793	0.5202	142	1.88	0.0618
rater	Rose		-0.08430	0.4930	142	-0.17	0.8645

The random effect estimates make sense. For example, 1 is administrative, and the random effect estimate is below average…we would not expect administrative personnel to have high exposure to burn pits. However, we might expect Infantry (Mos_num=3) to have higher exposure to burn pits, which the raters also conclude.

We see that Pepper scores job types higher, on average, with respect to Burn pit exposure, compared with the average rater; similarly, Gottschall scores lower. These both occur with marginal significance.

From our ordinal logistic regression model, we note that $P(Y_{ij} \geq k \mid b_i,\ b_j) = \frac{1}{1+e^{-\lambda_{ijk}}}$ and $P(Y_{ij} \geq k | b_i = 0,\ b_j = 0) = \frac{1}{1+e^{-\alpha_k}}$. From the latter, we can estimate that for an average rater and MOS_num, the probability of 'high' classification is $\frac{1}{1+e^{3.8512}} = 0.02$.

Job and rater-specific probability estimates can be obtained by using the first formula. We can also compute for specific MOS_num or raters, holding the other at its mean, since random effects are crossed. For example, for an average MOS_num the probability of a high classification for Gottschall is $\frac{1}{1+e^{-(-3.8512-1.15)}} = 0.7\%$, while for Pepper it is $\frac{1}{1+e^{-(-3.85+0.98)}} = 5.4\%$.

We can get probabilities for any given level by computing the cumulative probabilities, and then taking differences [e.g., $P(Y = 2) = P(Y \geq 2) - P(Y \geq 3)$.]

Comparing the descriptive and modeled approaches

Going back to the probabilities and odds we determined for the descriptive approach, why do they differ from the modeled approach?
Descriptive: $P(M \text{ or } H) = 37.5\%$ $Odds(M \text{ or } H) = 0.6$
$P(H) = 7.1\%$ $Odds(H) = 0.076$
Modeled: $P(M \text{ or } H) = 31.3\%$ $Odds(M \text{ or } H) = 0.46$ $P(H) = 2\%$
$Odds(H) = 0.02$

Why the difference? It appears that taking the correlation into account affects results; if the random effects are removed, here is what we get from the model (same as descriptive, above): Modeled:
$P(M \text{ or } H) = 37.5\%$ $Odds(M \text{ or } H) = 0.6$
$P(H) = 7.1\%$ $Odds(H) = 0.076$

Using the mixed-effects ordinal logistic regression for longitudinal data

We can generalize the formula for the mixed-effects ordinal logistic regression model so that it can be used for clustered / longitudinal data and include covariates. One such model that is useful for repeated measures within subjects (or subjects within clusters) is

$$\lambda_{ijk} = log\left[\frac{P(Y_{ij} \leq k \mid \boldsymbol{b}_i)}{1-P(Y_{ij} \leq k | \boldsymbol{b}_i)}\right] = \alpha_k + \boldsymbol{x}_{ij}^r \boldsymbol{\beta} + \boldsymbol{z}_{ij}^r \boldsymbol{b}_i$$

where i denotes subject, with measure j (or subject j in cluster i). Here, we have hierarchical data and so the random effects (as is usually done) are defined for the level 2 data (subjects).

The previous model can be used for longitudinal ordinal logistic regression, although we only account for repeated measures via random effects. (Using pseudo-likelihood methods, you could consider models that account for random effects or serial correlation, or both.)

Now we have what is called a proportional odds model (see McCullagh, 1980) that results from the fact that the relationship between the cumulative logit and the predictors does not depend on k.

For example, say that the previous case study also had measurements over time (x=time). If we added this as a predictor, then the cumulative logits (and hence probabilities) would not change over time.

We can generalize the model slightly so that for certain predictors, we do not require the proportional odds assumption.

For example, Hedeker and Mermelstein (1998, 200) suggest the model $\lambda_{ijk} = log\left[\frac{P(Y_{ij} \leq k \mid \boldsymbol{b}_i)}{1-P(Y_{ij} \leq k \mid \boldsymbol{b}_i)}\right] = \alpha_k + \boldsymbol{x}_{ij}^r\boldsymbol{\beta} + \boldsymbol{s}_{ij}^r\boldsymbol{\gamma}_k + \boldsymbol{z}_{ij}^r\boldsymbol{b}_i$, where the additional term involving γ_k allows the effects for the associated covariates to vary across the cumulative logits.

For more detail, see the above references or Hedeker and Gibbons (2006). Hedeker does warn about use of this partial proportional odds model, with respect to inference for certain values of the covariates. For more detail, see Hedeker and Gibbons (2006).

Summary