Topic 7: Analysis of Multi-category Outcomes

Nov22

Questions to be solved

1. pp 105 large sample size? Score Test is not reliable
2. Pp106 preference? We do prefer proportional model since it's simpler

Logit Models for Nominal Responses

1. **Categories**: J choose 2 in total, J - 1 non-redundant, others are redundant
   $log\frac{\pi_j(x)}{\pi_J(x)} = \alpha_j + \beta_j^T x, j = 1 ... J - 1$
2. **Baseline-Category Logits**:
   1. $log\frac{\pi_a(x)}{\pi_b(x)} = log\frac{\pi_a(x)}{\pi_J(x)} - log\frac{\pi_b(x)}{\pi_J(x)}$
   2. $\alpha_a - \alpha_b + (\beta_a - \beta_b)x$
   3. each **comparison** has its own intercepts and beta coefficient, J = 2: ordinary logistic regression for binary responses
3. **Alligator food choice example**
   1. $log(\pi_F/\pi_O) = 1.618 - 0.11x$
   2. $log(\pi_I/\pi_O) = 5.697 - 2.465x$
4. **odds ratio** = e^{β} estimated odds = e^{β} times the estimated odds at length x
   - eg. for alligators of length x+1 the estimated odds that the primary food type is "fish" rather than "other" equals 0.9 times the estimated odds at length x
5. **SAS likelihood ratio** Chi-square: 16.8, df = 2 , pf = 0.002 **H0 :**$\beta_f = \beta_I = 0$
6. **Estimated probability**: $\frac{e^{\alpha_j + \beta_j x}}{\sum_h e^{\alpha_h + \beta_h x}}, j = 1 ... J$
7. Contigency tables having many cells with small counts are said to be **Sparse**
   1. X^2 and G^2 better for comparing models than for testing model fit(for not sparse).
   2. $G^2(Model) = 2\sum observed[log\frac{observed}{fitted}]$
   3. $X^2(Model) = \sum \frac{(oberseved - fitted)^2}{fitted}$
   4. **Larger values of G2 and X2 provide evidence of lack of fit**, when fitted count are all >5, follows χ^2_{df}, df residual df = # parameters in saturated model(# of setting of the predictors: number of binomial observations for the data) - # parameters in the model of interest
   5. Logistics Regression with continuous predictor, X and G stat do not have approx chi-square

distribution.

8. $\pi_j(x) = \frac{exp(\alpha_j + \beta_j' x)}{1 + \sum EXP}$
9. **Binomial VS multinomial logistic Regression**: Effect of x on one category may not be monotonic
10. β estimate changes by 10% or siginificant , keep /add to the model

Logit Models for Odered(ordinal) Responses

1. **Ordinal** response: proportional odds model, or cumulative logit model
2. Response variable with a natural order: The ordinal nature of the outcome.
3. **Cumulative Logit**: $logit[P(Y \leq j)] = log[P(Y \leq j)/(1 - P(Y \leq j))] = log[\pi_1 ... + \pi_j/(\pi_{j+1} + ... + \pi_J)]$
4. **Proportional odds model**: provides a overall direction of a response and does not focus on specific outcome categories.
    1. $logit[P(Y \leq j)] = \alpha j + \beta x, j = 1, ..., J - 1$ beta fixed, pnly one beta
    2. Model **assumes** the effect of x is identical for all J-1 cumulative logits
    3. Assumption can be tested SCORE TEST
    4. Curves for each level of j have the same shape, shift by intercept
    5. Log of OR is $\beta(x_2 - x_1)$
5. **Proportional odds Assumption**: OR(Y<=1, x=2 vs x=1) == OR (Y<=2, x=2 vs x=1) == OR(Y<=3, x=2 vs x=1), it doesn't matter how you dichotomize the dependent var of the effects, the effects of the explanatory vars are the same
    1. Test if the assumption holds:
        1. OR(x1 vs x0) for π1 vs π2 and π3 = 435*(50+84)/((58+89)*275) = 1.44
        2. OR (x1 vs x0) for π1 and π2 vs π3 = ((435+58)*84)/(89*(275+50)) = 1.43
6. **Score Test**:
    1. high p value -> the assumption holds
    2. if more complex model fits significantly better, reject h0 at level 0.05, the assumption is not met, ie. OR's are different.
    3. Conservative for large sample size
    4. small p-value: separate binary logistic model VS cumulative logit model.---Prefer proportional odds model despite the jection of H0
7. **Assumption fails**:
    1. collapse two or more levels
    2. Add/remove additional predictors
    3. RUN Multinomial model
    4. Other ordinal models: partial proportional odds model, ajacent-category model, continuation-ratio model
8. **Logit models for ordered responses**:OR estimate: gender 1vs0 1.44, 95CI 1.09, 1.90
    1. female were more likely to report higher levels of spiritual beliefs than males

2. compared to males the odds of being in a lower category of belief rather than in a higher categories are 44%higher for females

9. Nomial: $\pi_1 vs \pi_3 = OR = 1.49$: the odds of believing versus not beliving in afterlife was 49%higher for males than for males

10. **Predicted value**: J = 4

    1. $\pi_3 = P(Y = 3) = P(Y \leq 3) - P(Y \leq 2)$
    2. $\pi_4 = P(Y = 4) = 1 - P(Y \leq 3)$