Xuange Liang
xl3493

Homework 6

1. Ordinal logistic regression model

a. Write down the model

$$logit[P(Y \leq k)] = \alpha_k + \beta_1 \times sex + \beta_2 \times age_2 + \beta_3 \times age_3$$

Where $k = 1$ (No\Little), 2(Important), Y is the ordinal response variable, α_k are the intercepts for each cumulative logit, β_1 is the coefficient for sex, β_2 β_3 are coefficients for age.

b. Test the proportional odds assumption

H_0: the proportional odds assumption holds (ordinal model is appropriate)
H_1: the proportional odds assumption doesn't hold

Score Test for the Proportional Odds Assumption		
Chi-Square	DF	Pr > ChiSq
0.7139	3	0.8699

Test statistics: Score test (Chi-Square) χ^2 =0.7139

Degrees of freedom: 3

P-value: 0.8699

Conclusion: We fail to reject the null hypothesis at α = 0.05. The proportional odds assumption is reasonable.

c. Estimate odds ratio and CI

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1	0.0433	0.2303	0.0354	0.8508
Intercept	2	1	1.6546	0.2536	42.5742	<.0001
sex	2	1	0.5762	0.2261	6.4936	0.0108
age	2	1	-1.1468	0.2773	17.1079	<.0001
age	3	1	-2.2322	0.2904	59.0806	<.0001

Odds ratio:

$$e^{0.5762} = 1.7793$$

95% CI:

$$CI_{lower} = e^{0.5762-1.96\times0.2261} = 1.1421$$
$$CI_{upper} = e^{0.5762+1.96\times0.2261} = 2.772$$

Therefore, the 95% CI is (1.1421 , 2.772).

Conclusion: Men have 1.7793 times the odds of giving a lower rating compared to women, adjusting for age. Because the 95% CI (1.1421, 2.772) doesn't include 1, we are 95% confident to say that this is true. This means that women care more about air conditioning and power steering features compared to men.

d. Estimate the probability

$$P(Y \leq 2) = \frac{e^{1.6546}}{1 + e^{1.6546}} = 0.8396$$
$$P(Y = 3) = 1 - P(Y \leq 2) = 0.1604$$

Therefore, the probability of rating "Very important" regarding the features of air conditioning and power steering in cars for women aged 18-23 is 0.1604.

2. Multinomial logistic regression model

a. Write down the model

$$logit\left(\frac{P(Y = k)}{P(Y = 1)}\right) = \beta_{0k} + \beta_{1k} \times sex + \beta_{2k} \times age_2 + \beta_{3k} age_3$$

Where $k = 2, 3$ (Important or Very Important), Y is the ordinal response variable, β_{0k} are the intercepts for each cumulative logit, β_{1k} are the coefficients for sex, $\beta_2 k$ β_{3k} are coefficients for age.

b. Estimate odds ratio and CI

Analysis of Maximum Likelihood Estimates

Parameter		response	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		2	1	-0.5908	0.2840	4.3286	0.0375
Intercept		3	1	-1.0391	0.3305	9.8843	0.0017
sex	2	2	1	-0.3881	0.3005	1.6677	0.1966
sex	2	3	1	-0.8129	0.3210	6.4122	0.0113
age	2	2	1	1.1283	0.3416	10.9059	0.0010
age	2	3	1	1.4780	0.4009	13.5912	0.0002
age	3	2	1	1.5876	0.4029	15.5270	<.0001
age	3	3	1	2.9165	0.4229	47.5594	<.0001

$$OR = e^{-0.813} = 0.4435$$

$$CI = e^{-0.813 \pm 1.96 \times 0.321} = (0.2364, 0.8321)$$

c. Estimate the probability

$$P = \frac{e^{-1.0391}}{1 + e^{-1.0391} + e^{-0.5908}} = 0.1854$$

Therefore, the probability of rating "Very important" regarding the features of air conditioning and power steering in cars for women aged 18-23 is 0.1854.

Xuange Liang
xl3493

3. Choosing model

I will choice the ordinal model because the proportional odds assumption is satisfied, and it is more simple. Also, since the response variable (importance rating) has a natural ordering, the ordinal model respects this structure and provides a more interpretable cumulative odds ratio across all levels.