Assignment 2

Milan Kuzmanovic, Mark McMahon
Martin Kotuliak, Jakub Polak

April 11, 2020

Task 1

The following table contains the estimates of a logistic regression model.

	Est.	s.e.	z	p-value	OR	95% C.I. for OR lower	higher
X_1	-0.868	**0.367**	-2.365	0.018	**0.42**	0.205	0.865
X_2	2.404	0.601	4.000	<0.001	**11.067**	**3.408**	**35.943**
X_3	**-3.604**	**0.511**	**-7.059**	<0.001	**0.027**	0.010	0.074

Fill in the missing information (Please report formulas and computation.)

$$\text{s.e.}_{X_1} = \frac{X_1}{z_{X_1}} = \frac{-0.868}{-2.365} = 0.367$$

$$\text{OR}_{X_1} = e^{X_1} = e^{-0.868} = 0.42$$

$$\text{OR}_{X_2} = e^{X_2} = e^{2.404} = 11.067$$

$$\text{lower}_{X_2} = e^{X_2 - z_{0.975} \times \text{s.e.}_{X_2}} = e^{2.404 - 1.96 \times 0.601} = 3.408$$

$$\text{lower}_{X_2} = e^{X_2 + z_{0.975} \times \text{s.e.}_{X_2}} = e^{2.404 + 1.96 \times 0.601} = 35.943$$

$$X_3 = \frac{\ln \text{lower}_{X_3} + \ln \text{higher}_{X_3}}{2} = \frac{0.01 + \ln 0.074}{2} = -3.604$$

$$\text{s.e.}_{X_3} = \frac{-\ln \text{lower}_{X_3} + \ln \text{higher}_{X_3}}{2 \times z_{0.975}} = \frac{-0.01 + \ln 0.074}{2 \times 1.96} = 0.511$$

$$z_{X_3} = \frac{X_3}{\text{s.e.}_{X_3}} = \frac{-3.604}{0.511} = -7.059$$

$$\text{OR}_{X_3} = e^{X_3} = e^{-3.604} = 0.027$$

Task 2

During the lecture, we have considered three systems of hypotheses for the parameters of the MNRM:

1. $H_0 : \beta_{jm} = 0$ vs. $H_1 : \beta_{jm} \neq 0$

2. $H_0 : \beta_{j1} = ... = \beta_{j(M-1)} = 0$ vs. $H_1 :$ at least one $\beta_{jm} \neq 0$, $\forall m$

3. $H_0 : \beta_{j1} = ... = \beta_{j(M-1)} = 0$ vs. $H_1 :$ at least one $\beta_{jm} \neq 0$, $\forall m, j$

Could you specify another pair of hypotheses H_0 and H_1 for the parameters of the MNRM that we might want to test? Justify your answer.

Firstly, it is important to explain which hypotheses can be tested, i.e. for which hypotheses is there a procedure to define a valid (asymptotic) test. The above hypotheses are tested either based on the asymptotic normality of the maximum likelihood estimator, or on the likelihood ratio statistic whose logarithm is asymptotically Chi-square distributed. It is important to notice that we can extend the hypothesis test $H_0 : \beta_{jm} = 0$ vs. $H_1 : \beta_{jm} \neq 0$ that is based on the asymptotic normality of the ML estimator to a joint test for any set of parameters. We can do this because the ML estimator is jointly normally distributed, so any subset of the set of all parameters is also jointly normally distributed, which means that by taking the quadratic form of the estimator minus hypothesized value and covariance matrix estimated by the inverse of the Fisher information matrix, we can define a Chi-square test statistic for any null hypotheses that an arbitrary subset of parameters is simultaneously equal to some specific hypothesized values (generally to zero). Therefore, we can in general construct a test for any joint hypothesis on the parameters. Now comes the question of which hypotheses might be of interest, i.e. which questions we might want an answer to? We propose several possible questions of interest defined by the below hypotheses:

1. $H_0 : \beta_{jm} = c$ vs. $H_1 : \beta_{jm} \neq c$ where $c \in \mathbb{R}$ is some constant. Here we have the same logic of the test as for $c = 0$. In fact, this is just a generalization that is useful if we want to answer more specific questions about the parameters. We might be interested in testing not just for the existence of a significant effect ($c = 0$ case), but also about specific magnitude of the effect (e.g. $c = 1$).

2. $H_0 : \beta_{jm} = 0$ vs. $H_1 : \beta_{jm} > 0$. We can change the alternative hypothesis in case we want to test whether there is a positive effect. Here, we assume that the parameter can't be negative, only zero under the null hypothesis or positive under the alternative hypothesis. This version of the test with different, so-called one-sided alternative, gives us more power in detecting positive (or negative in case $H_1 : \beta_{jm} < 0$) effects because we assume that the effect is either zero or positive, so the rejection region for given significance α is two times larger on the positive (or negative) side compared to the case with the two-sided alternative.

3. $H_0 : \beta_{j1} = ... = \beta_{j(M-1)} = c$ vs. $H_1 :$ at least one $\beta_{jm} \neq c$, $\forall m$, where $c \in \mathbb{R}$ is some constant. This is a variation of the second hypothesis test that was introduced in the lecture. Instead of testing if all of the parameters associated with variable j are zero, which would mean that variable j has no significant influence on the outcome variable, we can test that all of the parameters associated with variable j have the same effect $c \in \mathbb{R}$, on the odds relative to the reference group M. For example, having three groups, and group 3 as the reference, we might be interested if the increase in variable j results in the same change in the odds for group 1 relative to group 3 and for group 2 relative to group 3.

4. $H_0 : \beta_{1m} = ... = \beta_{pm} = 0$ vs. $H_1 :$ at least one $\beta_{jm} \neq 0$, $\forall j$. This hypothesis test would test if the variables in the model have any influence on the odds between group m and reference group M. We might be interested in this question if we suspect that for a specific group m the variables in the model don't affect the odds of m happening relative to M happening.

5. $H_0 : \beta_{j_1 m} - \beta_{j_2 m} = 0$ vs. $H_1 : \beta_{j_1 m} - \beta_{j_2 m} \neq 0$ This hypothesis test would test if the effect of the variable j_1 and the variable j_2 on the odds of group m relative to reference group M is the same.

There are many more possibilities for hypotheses tests. So, if we have a question of interest we can define a hypothesis test for that question, and likelihood theory allows us to construct asymptotic pivots for testing all sorts of hypotheses and thus answering all sorts of questions of interest.