

## Solutions for Examination Categorical Data Analysis, February 14, 2020

### Problem 1

- a. The linear logistic regression model has

$$\pi(x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)}. \quad (1)$$

- b. We need to find a confidence interval for  $\pi(8)$ . We first look at  $\text{logit}[\pi(8)] = \alpha + 8\beta$ , whose point estimate is

$$\text{logit}[\hat{\pi}(8)] = \hat{\alpha} + 8\hat{\beta} = -6.0 + 8 \cdot 0.5 = -2.0. \quad (2)$$

Since

$$\begin{aligned} \text{Var}[\text{logit}(\hat{\pi}(8))] &= \text{Var}(\hat{\alpha}) + 2 \cdot 8 \cdot \text{Cov}(\hat{\alpha}, \hat{\beta}) + 8^2 \cdot \text{Var}(\hat{\beta}) \\ &= \text{Var}(\hat{\alpha}) + 16 \cdot \text{Cov}(\hat{\alpha}, \hat{\beta}) + 64 \cdot \text{Var}(\hat{\beta}), \end{aligned}$$

this gives a standard error for the estimate in (2) that equals

$$\begin{aligned} \text{SE} &= \sqrt{\widehat{\text{Var}}(\hat{\alpha}) + 16 \cdot \widehat{\text{Cov}}(\hat{\alpha}, \hat{\beta}) + 64 \cdot \widehat{\text{Var}}(\hat{\beta})} \\ &= \sqrt{0.1 + 16 \cdot (-0.01) + 64 \cdot 0.005} \\ &= \sqrt{0.260} \\ &= 0.510 \end{aligned}$$

and a Wald type confidence interval

$$(-2.0 - 1.96 \cdot \text{SE}, -2.0 + 1.96 \cdot \text{SE}) = (-3.000, -1.001)$$

for  $\text{logit}[\pi(8)]$  with approximate coverage probability 95%, since  $z_{0.025} = \sqrt{\chi_1^2(0.05)} = 1.96$  is the 97.5% quantile of a standard normal distribution. The corresponding confidence interval for  $\pi(8)$ , with approximate coverage probability 95%, is

$$\left( \frac{\exp(-3.000)}{1 + \exp(-3.000)}, \frac{\exp(-1.001)}{1 + \exp(-1.001)} \right) = (0.048, 0.269).$$

- c. Suppose Scott's PSA level is  $x$ , so that James' PSA level is  $x + 3$ . The requested odds ratio is

$$\text{OR} = \frac{\pi(x+3)/(1-\pi(x+3))}{\pi(x)/(1-\pi(x))} = \frac{\exp(\alpha + \beta(x+3))}{\exp(\alpha + \beta x)} = \exp(3\beta).$$

By a similar argument as in a), we first compute a confidence interval

$$\begin{aligned} \left( \hat{\beta} - 1.96\sqrt{\widehat{\text{Var}}(\hat{\beta})}, \hat{\beta} + 1.96\sqrt{\widehat{\text{Var}}(\hat{\beta})} \right) &= (0.5 - 1.96 \cdot \sqrt{0.005}, 0.5 + 1.96 \cdot \sqrt{0.005}) \\ &= (0.3614, 0.6386) \end{aligned}$$

for  $\beta$  with approximate coverage probability 95%. The corresponding confidence interval for the odds ratio is

$$(\exp(3 \cdot 0.3614), \exp(3 \cdot 0.6386)) = (2.96, 6.79).$$

## Problem 2

- a. Let  $X, Y \in \{1, 2\}$  refer to type of surgery and outcome of surgery respectively, for a randomly chosen individual, and  $\pi_{ij} = P(X = i, Y = j)$  the probability that an observation belongs to cell  $(i, j)$ . Since these probabilities are proportional to the expected cell counts  $\mu_{ij}$  and sum to 1, we have that  $\pi_{ij} = \mu_{ij}/\mu_{++}$ . Therefore,

$$\begin{aligned} \pi_1 &= P(Y = 1|X = 1) = \pi_{11}/(\pi_{11} + \pi_{12}) = \mu_{11}/(\mu_{11} + \mu_{12}), \\ \pi_2 &= P(Y = 1|X = 2) = \pi_{21}/(\pi_{21} + \pi_{22}) = \mu_{21}/(\mu_{21} + \mu_{22}). \end{aligned}$$

- b. If one conditions on the two row sums  $N_{1+} = n_{1+} = n_1$  and  $N_{2+} = n_{2+} = n_2$ , we get independent binomial rows sampling. This corresponds to  $N_{11}$  and  $N_{21}$  being independent binomially distributed random variables

$$\begin{aligned} N_{11} &\sim \text{Bin}(n_1, \pi_1), \\ N_{21} &\sim \text{Bin}(n_2, \pi_2). \end{aligned}$$

- c. The maximum likelihood estimator of  $r$  is

$$\hat{r} = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{N_{11}/n_1}{N_{21}/n_2} = \frac{30/200}{20/100} = 0.75.$$

By means of a first order Taylor expansion, we have that

$$\begin{aligned} \log(\hat{r}) &= \log(\hat{\pi}_1) - \log(\hat{\pi}_2) \\ &\approx \log(\pi_1) + \frac{\hat{\pi}_1 - \pi_1}{\pi_1} - \log(\pi_2) - \frac{\hat{\pi}_2 - \pi_2}{\pi_2}. \end{aligned}$$

Consequently, since  $\hat{\pi}_1$  and  $\hat{\pi}_2$  are independent for the sampling scheme in b), the variance of  $\log(\hat{r})$  satisfies

$$\begin{aligned} \text{Var}[\log(\hat{r})] &\approx \text{Var}\left(\frac{\hat{\pi}_1 - \pi_1}{\pi_1}\right) + \text{Var}\left(\frac{\hat{\pi}_2 - \pi_2}{\pi_2}\right) \\ &= \frac{\text{Var}(\hat{\pi}_1)}{\pi_1^2} + \frac{\text{Var}(\hat{\pi}_2)}{\pi_2^2} \\ &= \frac{\text{Var}(N_{11})}{n_1^2 \pi_1^2} + \frac{\text{Var}(N_{21})}{n_2^2 \pi_2^2} \\ &= \frac{n_1 \pi_1 (1 - \pi_1)}{n_1^2 \pi_1^2} + \frac{n_2 \pi_2 (1 - \pi_2)}{n_2^2 \pi_2^2} \\ &= \frac{\pi_1 (1 - \pi_1)}{n_1 \pi_1^2} + \frac{\pi_2 (1 - \pi_2)}{n_2 \pi_2^2} \\ &= \frac{1 - \pi_1}{n_1 \pi_1} + \frac{1 - \pi_2}{n_2 \pi_2}. \end{aligned} \tag{3}$$

- d. The standard error of  $\log(\hat{r})$  is obtained by plugging in estimates of  $\pi_1$  and  $\pi_2$  into the variance formula (3) and then taking the square root. That is,

$$\begin{aligned}
\text{SE}[\log(\hat{r})] &= \sqrt{\widehat{\text{Var}}[\log(\hat{r})]} \\
&= \sqrt{\frac{1-\hat{\pi}_1}{n_1\hat{\pi}_1} + \frac{1-\hat{\pi}_2}{n_2\hat{\pi}_2}} \\
&= \sqrt{\frac{n_{12}}{n_1n_{11}} + \frac{n_{22}}{n_2n_{21}}} \\
&= \sqrt{\frac{170}{200 \cdot 30} + \frac{80}{100 \cdot 20}} \\
&= \sqrt{0.0683} \\
&= 0.2614.
\end{aligned}$$

This gives a Wald-based confidence interval

$$\begin{aligned}
&(\log(\hat{r}) - 1.96 \cdot \text{SE}[\log(\hat{r})], \log(\hat{r}) + 1.96 \cdot \text{SE}[\log(\hat{r})]) \\
&= (\log(0.75) - 1.96 \cdot 0.2614, \log(0.75) + 1.96 \cdot 0.2614) \\
&= (-0.800, 0.225)
\end{aligned}$$

for  $\log(r)$  with approximate coverage probability 95%. Applying the exponential transformation to both sides of this interval, we finally obtain a confidence interval

$$(\exp(-0.800), \exp(0.225)) = (0.45, 1.25)$$

for  $r$  with approximate coverage probability 95%. Since 1 belongs to this interval, the null hypothesis  $H_0$  that type of surgery does not influence the probability of outcome (i.e.  $\pi_1 = \pi_2$  or  $r = 1$ ) is not rejected.

## Problem 3

- a. For the loglinear model ( $AC, AM, CM$ ) we have that

$$\mu_{acm} = \exp(\lambda + \lambda_a^A + \lambda_c^C + \lambda_m^M + \lambda_{ac}^{AC} + \lambda_{am}^{AM} + \lambda_{cm}^{CM}),$$

for all cells  $a, c, m \in \{0, 1\}$ . If  $a = c = m = 0$  are chosen as baseline levels, then all loglinear parameters equal 0 if at least index is 0. This gives a parameter vector with the remaining nonzero loglinear parameters

$$\boldsymbol{\beta} = (\lambda, \lambda_1^A, \lambda_1^C, \lambda_1^M, \lambda_{11}^{AC}, \lambda_{11}^{AM}, \lambda_{11}^{CM}).$$

The number of parameters is thus  $p(AC, AM, CM) = 7$ .

- b. All of the listed models in the tables are balanced, and all three categorical variables are binary. Therefore, each model has 1 baseline parameter, 3 main effect parameters (1 per main effect),  $1 = (2-1) \cdot (2-1)$  parameter per second order association, and  $1 = (2-1) \cdot (2-1) \cdot (2-1)$  parameter per third order association. Adding the number of baseline, main effect, second order, and third order association parameters, we

find the total number of parameters

$$\begin{aligned}
p(A, C, M) &= 1 + 3 + 0 + 0 = 4, \\
p(A, CM) &= 1 + 3 + 1 + 0 = 5, \\
p(C, AM) &= 1 + 3 + 1 + 0 = 5, \\
p(M, AC) &= 1 + 3 + 1 + 0 = 5, \\
p(AC, AM) &= 1 + 3 + 2 + 0 = 6, \\
p(AC, CM) &= 1 + 3 + 2 + 0 = 6, \\
p(AM, CM) &= 1 + 3 + 2 + 0 = 6, \\
p(AC, AM, CM) &= 1 + 3 + 3 + 0 = 7, \\
p(ACM) &= 1 + 3 + 3 + 1 = 8
\end{aligned}$$

of all models.

- c. Since  $(ACM)$  is the saturated model, Akaike's Information Criterion of each model is

$$\begin{aligned}
\text{AIC}(\text{Model}) &= -2L(\text{Model}) + 2p(\text{Model}) \\
&= -2[L(\text{Model}) - L(ACM)] + 2p(\text{Model}) - 2L(ACM) \\
&= G^2(\text{Model}) + 2p(\text{Model}) - 2L(ACM),
\end{aligned}$$

where  $L(\text{Model})$  and  $G^2(\text{Model})$  are the log likelihood and deviance of each model respectively. We select the best model, according to the AIC-criterion, by minimizing  $\text{AIC}(\text{Model})$ , which is equivalent to minimizing  $G^2(\text{Model}) + 2p(\text{Model})$ . We found the number of parameters  $p(\text{Model})$  of all models in b). This makes it possible to fill in the second column of the given table, and then add a third column:

Model	$G^2$	$p$	$G^2 + 2p$
(A,C,M)	1286.0	4	1294.0
(A,CM)	534.2	5	544.2
(C,AM)	939.6	5	949.6
(M,AC)	843.8	5	853.8
(AC,AM)	497.4	6	509.4
(AC,CM)	92.0	6	104.0
(AM,CM)	187.8	6	199.8
(AC,AM,CM)	0.4	7	14.4
(ACM)	0.0	8	16

Since  $(AC, AM, CM)$  minimizes  $G^2(\text{Model}) + 2p(\text{Model})$ , this is the model chosen by the AIC-criterion.

- d. In the first step of backward elimination (BE), the largest model among those listed in the table,  $(ACM)$ , is tested against the model  $(AC, AM, CM)$  for which the third order association between the three variables has been removed, by means of a likelihood ratio test. This gives

$$\begin{aligned}
G^2(AC, AM, CM|ACM) &= -2[L(AC, AM, CM) - L(ACM)] \\
&= G^2(AC, AM, CM) - G^2(ACM) \\
&= 0.4 - 0 \\
&< \chi_{8-7}^2(0.05) = 3.84,
\end{aligned} \tag{4}$$

where  $df = p(ACM) - p(AC, AM, CM) = 8 - 7 = 1$  is used for the quantile of the  $\chi^2$ -distribution. Since the deviance does not exceed this quantile, the null hypothesis  $(AC, AM, CM)$  is not rejected, and the smaller model is selected in favor of the alternative hypothesis that  $(ACM)$  holds but not  $(AC, AM, CM)$ .

In the second step of the BE scheme we test  $(AC, AM, CM)$  against each one of the three models obtained by removing one second order association from  $(AC, AM, CM)$ . The log likelihood ratios of these three tests are found, as in (4), from the difference in deviance:

$$\begin{aligned} G^2(AC, AM|AC, AM, CM) &= 497.4 - 0.4 = 497.0, \\ G^2(AC, CM|AC, AM, CM) &= 92.0 - 0.4 = 91.6, \\ G^2(AM, CM|AC, AM, CM) &= 187.8 - 0.4 = 187.4, \end{aligned}$$

Since all these three deviances (by a very large margin) exceed  $\chi_{7-6}^2(0.05) = 3.84$ , the null hypothesis (the smaller model) is rejected in each test. Therefore, the BE scheme stops after this second step and  $(AC, AM, CM)$  is selected, the same model that was chosen with the AIC criterion in c).

## Problem 4

- a. Let  $\pi_{acm} = \mu_{acm}/\mu_{+++}$  be the probability of cell  $(a, c, m)$  for multinomial sampling when we condition on the total number of observations of the Poisson model  $(AC, AM, CM)$ . Regarding  $M$  as the outcome variable and  $A, C$  as predictor variables of this multinomial model, we find that  $M|A, C$  is an ANOVA type logistic regression model, since

$$\begin{aligned} \text{logit}P(M = 1|A = a, C = c) &= \log[P(M = 1|A = a, C = c)/P(M = 0|A = a, C = c)] \\ &= \log[(\pi_{ac1}/\pi_{ac+})/(\pi_{ac0}/\pi_{ac+})] \\ &= \log(\pi_{ac1}/\pi_{ac0}) \\ &= \log(\mu_{ac1}/\mu_{ac0}) \\ &= \log(\mu_{ac1}) - \log(\mu_{ac0}) \\ &= \lambda + \lambda_a^A + \lambda_c^C + \lambda_1^M + \lambda_{ac}^{AC} + \lambda_{a1}^{AM} + \lambda_{c1}^{CM} \\ &\quad - (\lambda + \lambda_a^A + \lambda_c^C + \lambda_0^M + \lambda_{ac}^{AC} + \lambda_{a0}^{AM} + \lambda_{c0}^{CM}) \\ &= \alpha + \beta_a^A + \beta_c^C, \end{aligned} \tag{5}$$

with

$$\begin{aligned} \alpha &= \lambda_1^M - \lambda_0^M = \lambda_1^M, \\ \beta_a^A &= \lambda_{a1}^{AM} - \lambda_{a0}^{AM} = \lambda_{a1}^{AM}, \\ \beta_c^C &= \lambda_{c1}^{CM} - \lambda_{c0}^{CM} = \lambda_{c1}^{CM}. \end{aligned}$$

In the last step we assumed that  $a = c = m = 0$  are baseline levels, putting to zero all loglinear parameters with at least one 0 index. Then all effect parameters  $\beta_0^A = \beta_0^C = 0$  vanish, and the remaining three nonzero parameters of the logistic regression model, are  $\boldsymbol{\beta} = (\alpha, \beta_1^M, \beta_1^C)$ .

- b. The conditional odds ratio of marijuana use between those that use alcohol and those that do not, conditional on cigarette use, is

$$\theta_{AM(c)} = \frac{P(M = 1|A = 1, C = c)/P(M = 0|A = 1, C = c)}{P(M = 1|A = 0, C = c)/P(M = 0|A = 0, C = c)}. \quad (6)$$

It follows from (5) that

$$\begin{aligned} \log \theta_{AM(c)} &= \text{logit}P(M = 1|A = 1, C = c) - \text{logit}P(M = 1|A = 0, C = c) \\ &= \alpha + \beta_1^A + \beta_c^C - (\alpha + \beta_0^A + \beta_c^C) \\ &= \beta_1^A - \beta_0^A \\ &= \beta_1^A \\ &= \lambda_{11}^{AM} \end{aligned}$$

when  $a = m = 0$  are chosen as baseline levels of alcohol and marijuana use. Equivalently,

$$\theta_{AM(c)} = \exp(\lambda_{11}^{AM}). \quad (7)$$

- c. There is homogeneous association between alcohol use  $A$  and marijuana use  $M$  if the conditional odds ratio  $\theta_{AM(c)}$  does not depend on the level  $c$  of smoking. It follows from (7) that model  $(AC, AM, CM)$  has homogeneous association, since the right hand side of this equation does not depend on  $c$ . Similarly, one shows that all loglinear models for which  $A$  and  $M$  are not involved in the same third order association, have homogeneous association between  $A$  and  $M$ . Hence, among the loglinear models listed in the table of Problem 3, all models except the saturated model  $(ACM)$  have homogeneous association between  $A$  and  $M$ .
- d. Since  $M$  and  $C$  are conditionally independent given  $A$  for model  $(AM, AC)$ , it follows that  $P(M|A, C) = P(M|A)$ . Inserting this relation into (6), we find that the conditional odds ratio

$$\theta_{AM(c)} = \frac{P(M = 1|A = 1)/P(M = 0|A = 1)}{P(M = 1|A = 0)/P(M = 0|A = 0)} = \theta_{AM} \quad (8)$$

for model  $(AM, AC)$  equals the marginal odds ratio  $\theta_{AM}$ , by the definition of the latter. We estimate the marginal odds ratio from the marginal twoway table of  $A, M$ , by replacing each  $P(M = m|A = a)$  in (8) with  $\hat{P}(M = m|A = a) = n_{a+m}/n_{a++}$ . Since all  $n_{0++}$  and  $n_{1++}$  cancel out, it follows that

$$\hat{\theta}_{AM} = \frac{n_{0+0}n_{1+1}}{n_{0+1}n_{1+0}} = \frac{(43 + 279) \cdot (911 + 44)}{(3 + 2) \cdot (538 + 456)} = 61.9.$$

This is very different from the estimate  $\hat{\theta}_{AM(c)} = 19.8$  of the conditional odds ratio between  $A$  and  $M$  for model  $(AC, AM, CM)$ . However, we know from Problems 3c) and 3d) that  $(AM, AC)$  fits data much worse than  $(AC, AM, CM)$ . For this reason the estimated (conditional) odds ratio of model  $(AC, AM, CM)$  is more trustworthy. Therefore, the odds of using marijuana is about 20 times higher for those that use alcohol, compared to those that don't.

## Problem 5

- a. This is a loglinear model with  $t_i$  as an offset. Let  $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)^T$  be the parameter vector. Then the likelihood function is

$$l(\boldsymbol{\lambda}) = \prod_{i=0}^3 e^{-\mu_i} \frac{\mu_i^{y_i}}{y_i!},$$

and the log likelihood

$$\begin{aligned} L(\boldsymbol{\lambda}) &= \log l(\boldsymbol{\lambda}) \\ &= \sum_{i=0}^3 [y_i \log(\mu_i) - \mu_i - \log(y_i!)] \\ &= \text{constant} + \sum_{i=0}^3 [y_i(\lambda_0 + \lambda_1 i) - t_i \exp(\lambda_0 + \lambda_1 i)], \end{aligned} \quad (9)$$

where

$$\text{constant} = \sum_{i=0}^3 [y_i \log(t_i) - \log(y_i!)]$$

does not depend on the two parameters  $\lambda_0$  and  $\lambda_1$ .

- b. Since

$$\mu_i = t_i \exp(\lambda_0 + \lambda_1 i), \quad (10)$$

we find that

$$\frac{d\mu_i}{d\boldsymbol{\lambda}} = \begin{pmatrix} \partial \mu_i / \partial \lambda_0 \\ \partial \mu_i / \partial \lambda_1 \end{pmatrix} = \mu_i \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

From this and (9) it follows that the likelihood score vector equals

$$\mathbf{u}(\boldsymbol{\lambda}) = \begin{pmatrix} \partial L(\boldsymbol{\lambda}) / \partial \lambda_0 \\ \partial L(\boldsymbol{\lambda}) / \partial \lambda_1 \end{pmatrix} = \sum_{i=0}^3 (y_i - \mu_i) \begin{pmatrix} 1 \\ i \end{pmatrix}. \quad (11)$$

The likelihood equations are obtained by solving

$$\mathbf{u}(\boldsymbol{\lambda})_{\boldsymbol{\lambda}=(\hat{\lambda}_0, \hat{\lambda}_1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

with respect to  $\hat{\lambda}_0$  and  $\hat{\lambda}_1$ , which is equivalent to solving

$$\sum_{i=0}^3 y_i \begin{pmatrix} 1 \\ i \end{pmatrix} = \sum_{i=0}^3 t_i \exp(\hat{\lambda}_0 + \hat{\lambda}_1 i) \begin{pmatrix} 1 \\ i \end{pmatrix}.$$

- c. We first find the Hessian matrix

$$\mathbf{H}(\boldsymbol{\lambda}) = \frac{d^2 L(\boldsymbol{\lambda})}{d^2 \boldsymbol{\lambda}} = \begin{pmatrix} \partial^2 L(\boldsymbol{\lambda}) / \partial^2 \lambda_0 & \partial^2 L(\boldsymbol{\lambda}) / (\partial \lambda_0 \partial \lambda_1) \\ \partial^2 L(\boldsymbol{\lambda}) / (\partial \lambda_0 \partial \lambda_1) & \partial^2 L(\boldsymbol{\lambda}) / \partial^2 \lambda_1 \end{pmatrix}$$

of the log likelihood by differentiating (11) with respect to  $\lambda_0$  and  $\lambda_1$ . This gives

$$\mathbf{H}(\boldsymbol{\lambda}) = - \sum_{i=0}^3 \mu_i \begin{pmatrix} 1 & i \\ i & i^2 \end{pmatrix}.$$

Since  $\mathbf{H}(\boldsymbol{\lambda})$  does not depend on data it is non-stochastic. Therefore the Fisher information matrix equals

$$\mathbf{J}(\boldsymbol{\lambda}) = -E[\mathbf{H}(\boldsymbol{\lambda})] = -\mathbf{H}(\boldsymbol{\lambda}) = \sum_{i=0}^3 \mu_i \begin{pmatrix} 1 & i \\ i & i^2 \end{pmatrix}. \quad (12)$$

- d. The covariance matrix  $\text{Cov}(\hat{\boldsymbol{\lambda}})$  of the parameter vector is approximately  $\mathbf{J}(\boldsymbol{\lambda})^{-1}$ . It is estimated by

$$\widehat{\text{Cov}}(\hat{\boldsymbol{\lambda}}) = \hat{\mathbf{J}}^{-1} = \begin{pmatrix} \hat{J}_{00}^{(-1)} & \hat{J}_{01}^{(-1)} \\ \hat{J}_{10}^{(-1)} & \hat{J}_{11}^{(-1)} \end{pmatrix},$$

where

$$\hat{\mathbf{J}} = \mathbf{J}(\hat{\boldsymbol{\lambda}}) = -\mathbf{H}(\hat{\boldsymbol{\lambda}}) = \sum_{i=0}^3 \hat{\mu}_i \begin{pmatrix} 1 & i \\ i & i^2 \end{pmatrix}.$$

is the observed (expected) Fisher information matrix, and  $\hat{\mu}_i = t_i \exp(\hat{\lambda}_0 + \hat{\lambda}_1 i)$  the estimated expected number of heart attacks within each group  $i$  of patients. The one-sided Wald test, with approximate significance level  $\alpha$ , rejects  $H_0$  in favor of  $H_a$  when

$$z_W = \frac{\hat{\lambda}_1}{\sqrt{\hat{J}_{11}^{(-1)}}}$$

exceeds the  $(1 - \alpha)$ -quantile  $z_\alpha$  of a standard normal distribution.