STOCKHOLM UNIVERSITY
DEPT OF MATHEMATICS
Div. of Mathematical statistics
Ola Hössjer

MT 5019
SOLUTIONS
February 20 2025

Solutions for Examination Categorical Data Analysis, February 20, 2025

Problem 1

a. The linear logistic regression model has

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

b. We want to find a 95% confidence interval for $\pi(1)$. In order to do so we first consider logit $[\pi(1)] = \alpha + \beta$, whose point estimate is

$$\text{logit}\,[\hat{\pi}(1)] = \hat{\alpha} + \hat{\beta} = -1.5 - 1.2 = -2.7. \tag{1}$$

Since

$$\text{Var}\,\{\text{logit}[\hat{\pi}(1)]\} = \text{Var}(\hat{\alpha}) + \text{Var}(\hat{\beta}) + 2\text{Cov}(\hat{\alpha}, \hat{\beta}),$$

the standard error of the estimate in (1) is

$$\begin{aligned} \text{SE} &= \sqrt{\widehat{\text{Var}}\,\{\text{logit}[\hat{\pi}(1)]\}} \\ &= \sqrt{\widehat{\text{Var}}(\hat{\alpha}) + \widehat{\text{Var}}(\hat{\beta}) + 2\widehat{\text{Cov}}(\hat{\alpha}, \hat{\beta})} \\ &= \sqrt{0.05 + 0.02 + 2 \cdot (-0.01)} \\ &= \sqrt{0.05}. \end{aligned}$$

This gives a confidence interval

$$(-2.7 - 1.96 \cdot \text{SE}, -2.7 + 1.96 \cdot \text{SE}) = (-3.1383, -2.2617)$$

for logit$[\pi(1)]$ with approximate coverage probability 95%, since $z_{0.025} = 1.96$ is the 97.5% quantile of a standard normal distribution. The corresponding confidence interval for $\pi(1)$, with approximate coverage probability 95%, is

$$\left(\frac{e^{-3.1383}}{1 + e^{-3.1383}}, \frac{e^{-2.2617}}{1 + e^{-2.2617}} \right) = (0.0416, 0.0943).$$

c. The probability of a new heart attack within five years is $\pi(1)$ for Ben and $\pi(2.5)$ for Josh. This gives an odds ratio

$$\text{OR} = \frac{\pi(1)/(1-\pi(1))}{\pi(2.5)/(1-\pi(2.5))} = \frac{\exp(\alpha+\beta)}{\exp(\alpha+2.5\beta)} = \exp(-1.5\beta),$$

and the accompanying maximum likelihood estimate is

$$\widehat{\text{OR}} = \exp(-1.5\hat{\beta}) = \exp\left[-1.5(-1.2)\right] = 6.05.$$

An approximate 95% confidence interval for β is

$$\begin{aligned} &\left(\hat{\beta} - 1.96 \cdot \sqrt{\widehat{\text{Var}}(\hat{\beta})}, \hat{\beta} + 1.96 \cdot \sqrt{\widehat{\text{Var}}(\hat{\beta})}\right) \\ &= (-1.2 - 1.96 \cdot \sqrt{0.02}, -1.2 + 1.96 \cdot \sqrt{0.02}) \\ &= (-1.4772, -0.9228), \end{aligned} \tag{2}$$

and the corresponding interval for the odds ratio is

$$I = (e^{-1.5(-0.9228)}, e^{-1.5(-1.4772)}) = (3.99, 9.17).$$

In the last step we used that $x \to e^{-1.5x}$ is a monotone decreasing function, so that the end points of the transformed interval I are switched compared to (2).

Problem 2

a. This is a loglinear model with n_i as an offset. We let $\boldsymbol{\lambda} = (\lambda_0, \lambda_1)^T$ refer to the parameter vector. 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

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

where

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

does not depend on the parameters λ_0 and λ_1.

b. Since

$$\mu_i = n_i \exp(\lambda_0 + \lambda_1 i), \tag{4}$$

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 (3) it follows that the likelihood score vector equals

$$\boldsymbol{u}(\boldsymbol{\lambda}) = \begin{pmatrix} u_0(\boldsymbol{\lambda}) \\ u_1(\boldsymbol{\lambda}) \end{pmatrix} = \begin{pmatrix} \partial L(\boldsymbol{\lambda})/\partial\lambda_0 \\ \partial L(\boldsymbol{\lambda})/\partial\lambda_1 \end{pmatrix} = \sum_{i=0}^{3}(y_i - \mu_i)\begin{pmatrix} 1 \\ i \end{pmatrix}. \tag{5}$$

The likelihood equations are obtained by solving

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

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

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

c. We first find the Hessian matrix

$$\boldsymbol{H}(\boldsymbol{\lambda}) = \frac{d^2L(\boldsymbol{\lambda})}{d^2\boldsymbol{\lambda}} = \begin{pmatrix} \partial^2 L(\boldsymbol{\lambda})/\partial^2\lambda_0 & \partial^2 L(\boldsymbol{\lambda})/(\partial\lambda_0\partial\lambda_1) \\ \partial^2 L(\boldsymbol{\lambda})/(\partial\lambda_0\partial\lambda_1) & \partial^2 L(\boldsymbol{\lambda})/\partial^2\lambda_1 \end{pmatrix} = \begin{pmatrix} \partial u_0(\boldsymbol{\lambda})/\partial\lambda_0 & \partial u_0(\boldsymbol{\lambda})/\partial\lambda_1 \\ \partial u_1(\boldsymbol{\lambda})/\partial\lambda_0 & \partial u_1(\boldsymbol{\lambda})/\partial\lambda_1 \end{pmatrix}$$

of the log likelihood by differentiating the score function components $u_0(\boldsymbol{\lambda})$ and $u_1(\boldsymbol{\lambda})$ in (5) with respect to λ_0 and λ_1. This gives

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

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

$$\boldsymbol{J}(\boldsymbol{\lambda}) = -E\left[\boldsymbol{H}(\boldsymbol{\lambda})\right] = -\boldsymbol{H}(\boldsymbol{\lambda}) = \sum_{i=0}^{3} \mu_i \begin{pmatrix} 1 & i \\ i & i^2 \end{pmatrix}. \tag{6}$$

d. By taking the logarithm of (4) for $i = 0, 1, 2, 3$, it follows that

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad \boldsymbol{c} = \begin{pmatrix} \log(n_0) \\ \log(n_1) \\ \log(n_2) \\ \log(n_3) \end{pmatrix}.$$

Combining this with (6), we find after some computations that

$$\boldsymbol{J}(\boldsymbol{\lambda}) = \boldsymbol{X}^T \begin{pmatrix} \mu_1 & 0 & 0 & 0 \\ 0 & \mu_2 & 0 & 0 \\ 0 & 0 & \mu_3 & 0 \\ 0 & 0 & 0 & \mu_4 \end{pmatrix} \boldsymbol{X}.$$

Problem 3

a. The loglinear model $M = (XY, YZ)$ has expected cell counts

$$\mu_{ijk} = \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}), \quad 0 \le i, j, k \le 1.$$

Some of its parameters must be put to zero in order for the others to be identifiable. If the lowest level 0 of each variable X, Y, Z is taken as a baseline, all parameters with a least one index at its lowest level are put to zero. The remaining six parameters are

$$\boldsymbol{\lambda} = (\lambda, \lambda_1^X, \lambda_1^Y, \lambda_1^Z, \lambda_{11}^{XY}, \lambda_{11}^{YZ}).$$

b. Let

$$\pi_{ijk} = \mu_{ijk}/\mu_{+++} \tag{7}$$

be the cell probabilities of the multinomial model obtained when conditioning on the total number of observations $N_{+++} = n_{+++}$ for model $M = (XY, YZ)$. Then introduce the conditional probabilities $\pi_{i|j} = P(X = i|Y = j)$ and $\pi_{k|j} = P(Z = k|Y = j)$. Since X and Z are conditionally independent given Y, it follows that

$$\pi_{ijk} = \pi_{+j+}\pi_{i|j}\pi_{k|j} = \pi_{+j+} \cdot \frac{\pi_{ij+}}{\pi_{+j+}} \cdot \frac{\pi_{+jk}}{\pi_{+j+}} = \frac{\pi_{ij+}\pi_{+jk}}{\pi_{+j+}}. \tag{8}$$

Inserting (8) into (7), we obtain the desired formula, since

$$\begin{aligned} \mu_{ijk} &= \mu_{+++}\pi_{ijk} \\ &= \mu_{+++}\frac{\pi_{ij+}\pi_{+jk}}{\pi_{+j+}} \\ &= \mu_{+++}\frac{(\mu_{ij+}/\mu_{+++})(\mu_{+jk}/\mu_{+++})}{\mu_{+j+}/\mu_{+++}} \\ &= \frac{\mu_{ij+}\mu_{+jk}}{\mu_{+j+}}. \end{aligned} \tag{9}$$

c. The fitted cell counts $\hat{\mu}_{ijk}$ for model M are obtained by replacing the expected cell counts on the right hand side of (9) by the observed ones, i.e.

$$\hat{\mu}_{ijk} = \frac{n_{ij+}n_{+jk}}{n_{+j+}}. \tag{10}$$

Starting with cell $(0, 0, 0)$, we read off the values from the full and marginal tables and find that

$$\hat{\mu}_{000} = \frac{n_{00+}n_{+00}}{n_{+0+}} = \frac{(154 + 20) \cdot 215}{250} = 149.64.$$

A similar calculation for the other cells gives

$$\begin{aligned} \hat{\mu}_{010} &= 108.04, \\ \hat{\mu}_{100} &= 65.36, \\ \hat{\mu}_{110} &= 76.97, \\ \hat{\mu}_{001} &= 24.36, \\ \hat{\mu}_{011} &= 37.96, \\ \hat{\mu}_{101} &= 10.64, \\ \hat{\mu}_{111} &= 27.04. \end{aligned}$$

d. Inserting the observed cell counts n_{ijk} from the contingency table and the fitted cell counts $\hat{\mu}_{ijk}$ from c) into the given formula for the deviance, we find that

$$\begin{aligned} G^2(M) &= 2\sum_{ijk} n_{ijk} \log \tfrac{n_{ijk}}{\hat{\mu}_{ijk}} \\ &= 2\Big[154 \log \tfrac{154}{149.64} + 116 \log \tfrac{116}{108.04} + 61 \log \tfrac{61}{65.36} + 69 \log \tfrac{69}{76.97} \\ &\quad +20 \log \tfrac{20}{24.36} + 30 \log \tfrac{30}{37.96} + 15 \log \tfrac{15}{10.64} + 35 \log \tfrac{35}{27.04}\Big] \\ &= 8.19. \end{aligned}$$

Since the deviance exceeds $\chi^2_2(0.05) = 5.99$, we reject model $M = (XY, YZ)$ at level 5%. In the last step we used that the saturated model has $2 \cdot 2 \cdot 2 = 8$ parameters, whereas in a) we found that M has 6 parameters. Therefore the number of degrees of freedom is $8 - 6 = 2$.

Problem 4

a. It follows that $Y|X, Z$ is a logistic type regression model, since

$$\begin{aligned} &\text{logit} P(Y = 1|X = i, Z = k) \\ &\quad = \log P(Y = 1|X = i, Z = k) - \log P(Y = 0|X = i, Z = k) \\ &\quad = \log(\pi_{i1k}/\pi_{i+k}) - \log(\pi_{i0k}/\pi_{i+k}) \\ &\quad = \log \pi_{i1k} - \log \pi_{i0k} \\ &\quad = \log(\mu_{i1k}/\mu_{+++}) - \log(\mu_{i0k}/\mu_{+++}) \\ &\quad = \log \mu_{i1k} - \log \mu_{i0k} \\ &\quad = (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{1k}^{YZ}) \\ &\quad = (\lambda + \lambda_i^X + \lambda_0^Y + \lambda_k^Z + \lambda_{i0}^{XY} + \lambda_{0k}^{YZ}) \\ &\quad = (\lambda_1^Y - \lambda_0^Y) + (\lambda_{i1}^{XY} - \lambda_{i0}^{XY}) + (\lambda_{1k}^{YZ} - \lambda_{0k}^{YZ}) \\ &\quad = \lambda_1^Y + \lambda_{i1}^{XY} + \lambda_{1k}^{YZ} \\ &\quad =: \alpha + \beta_i^X + \beta_k^Z, \end{aligned} \tag{11}$$

where in the second last step we assumed that $i = k = 0$ are chosen as baseline levels. Because of this, the nonzero parameters of the model are $\boldsymbol{\theta} = (\alpha, \beta_1^X, \beta_1^Z)$.

b. We deduce from equation (11) that

$$\begin{aligned} \log \theta_{ik} &= \text{logit} P(Y = 1|X = i, Z = k) - \text{logit} P(Y = 1|X = 0, Z = 0) \\ &= (\alpha + \beta_i^X + \beta_k^Z) - \alpha \\ &= \beta_i^X + \beta_k^Z. \end{aligned} \tag{12}$$

This implies

$$\begin{aligned} \theta_{01} &= \exp(\beta_1^Z), \\ \theta_{10} &= \exp(\beta_1^X), \\ \theta_{11} &= \exp(\beta_1^X + \beta_1^Z). \end{aligned} \tag{13}$$

c. We first rewrite the odds ratios as

$$\theta_{ik} = \frac{P(Y = 1|X = i, Z = k)/P(Y = 0|X = i, Z = k)}{P(Y = 1|X = 0, Z = 0)/P(Y = 0|X = 0, Z = 0)} \tag{14}$$

for $(i,k) \in \{(0,1),(1,0),(1,1)\}$. It follows from Bayes' Theorem that

$$P(Y=j|X=i,Z=k) = \frac{P(X=i,Z=k|Y=j)P(Y=j)}{P(X=i,Z=k)}. \tag{15}$$

Insert (15) into (14). We find after some simplifications that

$$\theta_{ik} = \frac{P(X=i,Z=k|Y=1)/P(X=i,Z=k|Y=0)}{P(X=0,Z=0|Y=1)/P(X=0,Z=0|Y=0)}, \tag{16}$$

since all the terms that involve the marginal distributions of Y and X,Z cancel out. It therefore follows from (16) that θ_{01}, θ_{10} and θ_{11} can all be expressed in terms of the $X,Z|Y$- distribution.

We know from a) that $\boldsymbol{\theta} = (\alpha, \beta_1^X, \beta_1^Z)$. Write $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta}_1^X, \hat{\beta}_1^Z)$ and let

$$\hat{\theta}_{ik} = \exp(\hat{\beta}_i^X + \hat{\beta}_k^Z) \to \theta_{ik}^* \tag{17}$$

be the estimate of θ_{ik} obtained from the maximum likelihood estimate, with asymptotic limit θ_{ik}^* as the number of cases and controls grows. Because of (16), this limit will only depend on the $X,Z|Y$-distribution that the sample is drawn from, which by the definition of a case-control study is identical to the population distribution of $X,Z|Y$. The odds ratios are therefore consistently estimated, i.e.

$$\theta_{ik}^* = \theta_{ik}, \quad (i,k) \in \{(0,1),(1,0),(1,1)\}. \tag{18}$$

We deduce from (13), (17) and (18) that the two effect parameters β_1^X and β_1^Z will be estimated consistently as well (whereas α will not).