# Homework 2

Denis Ostroushko

2023-02-16

## 1 Problem 1

**4.2** For Table 3.8 with scores (0, 0.5, 1.5, 4.0, 7.0) for alcohol consumption, ML fitting of the linear probability model for malformation has output:

| Parameter | Estimate | Std Error | Wald 95% Conf Limits | |
|---|---|---|---|---|
| Intercept | 0.0025 | 0.0003 | 0.0019 | 0.0032 |
| alcohol | 0.001087 | 0.000727 | -0.0003 | 0.0025 |

**a.** Interpret the model parameter estimates. Use the fit to estimate the relative risk of malformation that compares alcohol consumption levels 0 and 7.0.

**b.** Some software (such as R) also reports $\hat{\beta} = 0.001087$ but instead reports $SE = 0.000832$. Why do you think the $SE$ value is different? [*Hint*: The software with output shown above inverts the *observed* information matrix to obtain the standard errors.]

## 2 Problem 2

For binary data, define a GLM using log link (define any notation you use). Show that the effects refer to the relative risk. Why do you think this link is not often used? (Hint: what happens if the linear predictor takes a positive value?)

## 3 Problem 3

**5.2** For a study using logistic regression to determine characteristics associated with remission in cancer patients, Table 5.11 shows the most important explanatory variable, a labeling index (LI) that measures proliferative activity of cells after



a patient receives an injection of tritiated thymidine. It represents the percentage of cells that are "labeled." The response measured whether the patient achieved remission. Software reports Table 5.12 for a logistic regression model using LI to estimate $\pi = P(\text{remission})$.

**a.** Show how software obtained $\hat{\pi} = 0.068$ when LI = 8.

**b.** Show that $\hat{\pi} = 0.50$ when LI = 26.0.

**c.** Show that the rate of change in $\hat{\pi}$ is 0.009 when LI = 8 and 0.036 when LI = 26.

**d.** The lower quartile and upper quartile for LI are 14 and 28. Show that $\hat{\pi}$ increases by 0.42, from 0.15 to 0.57, between those values.

**e.** For a unit increase in LI, show that the estimated odds of remission multiply by 1.16.

**f.** Explain how to obtain the confidence interval reported for the odds ratio. Interpret.

**g.** Construct a Wald test for the effect. Interpret.

**h.** Conduct a likelihood-ratio test for the effect, showing how to construct the test statistic using the $-2 \log L$ values reported.

**i.** Show how software obtained the confidence interval for $\pi$ reported at LI = 8. [*Hint:* Use the reported covariance matrix.]

**Table 5.11 Data for Exercise 5.2 on Cancer Remission**

| LI | Number of Cases | Number of Remissions | LI | Number of Cases | Number of Remissions | LI | Number of Cases | Number of Remissions |
|---|---|---|---|---|---|---|---|---|
| 8 | 2 | 0 | 18 | 1 | 1 | 28 | 1 | 1 |
| 10 | 2 | 0 | 20 | 3 | 2 | 32 | 1 | 0 |
| 12 | 3 | 0 | 22 | 2 | 1 | 34 | 1 | 1 |
| 14 | 3 | 0 | 24 | 1 | 0 | 38 | 3 | 2 |
| 16 | 3 | 0 | 26 | 1 | 1 | | | |

*Source*: Data reprinted with permission from E. T. Lee, *Comput. Prog. Biomed.* **4**: 80–92, 1974.

**Table 5.12 Software Output (Based on SAS) for Exercise 5.2**

```
                    Intercept   Intercept and
      Criterion       Only        Covariates
      -2 Log L       34.372         26.073

Parameter   Estimate   Standard Error   Chi-Square   Pr > ChiSq
Intercept    -3.7771       1.3786          7.5064       0.0061
li            0.1449       0.0593          5.9594       0.0146

                  Odds Ratio Estimates
   Effect   Point Estimate   95% Wald Confidence Limits
   li            1.156         1.029           1.298

              Estimated Covariance Matrix
           Variable    Intercept       li
           Intercept    1.900616    -0.07653
           li          -0.07653      0.003521

     Obs   li   remiss   n    pi_hat     lower      upper
       1    8      0     2   0.06797   0.01121   0.31925
       2   10      0     2   0.08879   0.01809   0.34010
```

# 4 Problem 4

**5.14** Refer to the prediction equation $\text{logit}(\hat{\pi}) = -10.071 - 0.509c + 0.458x$ for model (5.14) using quantitative color and width. The means and standard deviations are $\bar{c} = 2.44$ and $s = 0.80$ for color, and $\bar{x} = 26.30$ and $s = 2.11$ for width. For standardized predictors [e.g., $x = (\text{width} - 26.30)/2.11$], explain why the estimated coefficients



of $c$ and $x$ equal $-0.41$ and $0.97$. Interpret these by comparing the partial effects of a standard deviation increase in each predictor on the odds. Describe the color effect by estimating the change in $\hat{\pi}$ between the first and last color categories at the sample mean width.

## 5 Problem 5

**5.18** In a study designed to evaluate whether an educational program makes sexually active adolescents more likely to obtain condoms, adolescents were randomly assigned to two experimental groups. The educational program, involving a lecture and videotape about transmission of HIV, was provided to one group but not the other. Table 5.17 summarizes results of a logistic regression model for factors observed to influence teenagers to obtain condoms.

**a.** Find the parameter estimates for the fitted model, using (1, 0) indicator variables for the first three predictors. Based on the corresponding confidence interval for the log odds ratio, determine the standard error for the group effect.

**b.** Explain why either the estimate of 1.38 for the odds ratio for gender or the corresponding confidence interval is incorrect. Show that if the reported interval is correct, 1.38 is actually the *log* odds ratio, and the estimated odds ratio equals 3.98.

**Table 5.17 Data for Exercise 5.18 on Obtaining Condoms**

| Variable | Odds Ratio | 95% Confidence Interval |
|---|---|---|
| Group (education vs. none) | 4.04 | (1.17, 13.9) |
| Gender (males vs. females) | 1.38 | (1.23.12.88) |
| SES (high vs. low) | 5.82 | (1.87, 18.28) |
| Lifetime number of partners | 3.22 | (1.08, 11.31) |

*Source*: V. I. Rickert et al., *Clin. Pediatr.* **31**: 205–210, 1992.

## 6 Problem 6

**5.34** Construct the log-likelihood function for the model $\text{logit}[\pi(x)] = \alpha + \beta x$ with independent binomial outcomes of $y_0$ successes in $n_0$ trials at $x = 0$ and $y_1$ successes in $n_1$ trials at $x = 1$. Derive the likelihood equations, and show that $\hat{\beta}$ is the sample log odds ratio.