Statistical Methods - Quiz #2(70 minutes)

April 8, 2026 (Wednesday)

Section(교반): __A1__ Cadet Number(교번): __________ Name(성명): ______________ Score: ____________

- All solutions must include a detailed step-by-step explanation.
- Reference sheet is available on the last page.

1. Consider the logistic regression model with a single binary covariate $x_i \in \{0, 1\}$ but without intercept:

$$\log \frac{p_i}{1-p_i} = \beta_1 x_i, \quad y_i \sim \text{Bernoulli}(p_i) \ \text{independently}, \quad i = 1, \cdots, n.$$

The data are summarized in the following contingency table:

		Y		
		0	**1**	**Total**
X	0	n_{00}	n_{01}	$n_{00} + n_{01} = n_0$
	1	n_{10}	n_{11}	$n_{10} + n_{11} = n_1$
	Total	$n_{00} + n_{10}$	$n_{01} + n_{11}$	$n_{00} + n_{01} + n_{10} + n_{11}$

(A) Show that

$$p_i = \frac{1}{1 + \exp(-\beta_1 x_i)}.$$

<u>Solution:</u> Exponentiating both sides of the logit equation gives $\frac{p_i}{1-p_i} = e^{\beta_1 x_i}$, so $p_i = (1 - p_i)e^{\beta_1 x_i}$. Solving for p_i:

$$p_i = \frac{e^{\beta_1 x_i}}{1 + e^{\beta_1 x_i}} = \boxed{\frac{1}{1 + \exp(-\beta_1 x_i)}}.$$

(B) For subjects with $x_i = 0$, $p_i = p(0) = \frac{1}{2}$; for $x_i = 1$, $p_i = p(1) = \frac{e^{\beta_1}}{1 + e^{\beta_1}}$. Express the odds ratio(OR)

$$\text{OR} = \frac{p(1)/(1 - p(1))}{p(0)/(1 - p(0))}$$

in terms of β_1.

<u>Solution:</u> From the logistic model, the odds for each group are

$$\frac{p(0)}{1 - p(0)} = 1, \qquad \frac{p(1)}{1 - p(1)} = e^{\beta_1}.$$

Therefore,

$$\text{OR} = \frac{p(1)/(1 - p(1))}{p(0)/(1 - p(0))} = \frac{e^{\beta_1}}{1} = \boxed{e^{\beta_1}.}$$

Hence $\beta_1 = \log(\text{OR})$ is the log odds ratio of $Y = 1$ for $X = 1$ versus $X = 0$.

(C) Show that

$$E(n_{00}) = E(n_{01}).$$

Hint: You may use $E(n_{00}) = E\left(\sum_{i=1}^n (1-y_i)(1-x_i)\right) = \sum_{i=1}^n P(y_i = 0 \mid x_i = 0)(1-x_i)$, and $E(n_{01}) = E\left(\sum_{i=1}^n y_i(1-x_i)\right) = \sum_{i=1}^n P(y_i = 1 \mid x_i = 0)(1-x_i)$.

Solution: From part (B), we have $p(0) = \dfrac{1}{2}$, so

$$P(y_i = 0 \mid x_i = 0) = 1 - p(0) = \frac{1}{2}, \qquad P(y_i = 1 \mid x_i = 0) = p(0) = \frac{1}{2}.$$

Substituting into the hint:

$$E(n_{00}) = \sum_{i=1}^n P(y_i = 0 \mid x_i = 0)(1-x_i) = \sum_{i=1}^n \frac{1}{2}(1-x_i) = \frac{n_0}{2},$$

$$E(n_{01}) = \sum_{i=1}^n P(y_i = 1 \mid x_i = 0)(1-x_i) = \sum_{i=1}^n \frac{1}{2}(1-x_i) = \frac{n_0}{2}.$$

Therefore $E(n_{00}) = E(n_{01}) = \dfrac{n_0}{2}$.

(D) Show that the closed form MLE(Maximum likelihood estimator) $\hat{\beta}_1$ that solves the score equation

$$S(\beta_1) = \sum_{i=1}^n (y_i - p_i)x_i = 0$$

is $\hat{\beta}_1 = \log \dfrac{n_{11}}{n_{10}}$

Solution: Since the model has no intercept, for $x_i = 0$ we have $\log \dfrac{p_i}{1-p_i} = 0$, so $p(0) = \frac{1}{2}$ regardless of β_1. Hence all terms with $x_i = 0$ vanish in the score equation, and the score reduces to

$$S(\beta_1) = \sum_{i:\, x_i = 1} \big(y_i - p(1)\big) = n_{11} - n_1 p(1) = 0,$$

giving $\hat{p}(1) = \dfrac{n_{11}}{n_1} = \dfrac{n_{11}}{n_{10} + n_{11}}$. Inverting the logistic link,

$$e^{\hat{\beta}_1} = \frac{\hat{p}(1)}{1 - \hat{p}(1)} = \frac{n_{11}/(n_{10}+n_{11})}{n_{10}/(n_{10}+n_{11})} = \frac{n_{11}}{n_{10}},$$

and therefore

$$\boxed{\hat{\beta}_1 = \log \frac{n_{11}}{n_{10}}.}$$

(E) The Fisher information is given as

$$\mathcal{I}(\beta_1) = \sum_{i=1}^{n} p_i(1-p_i)x_i^2 = \sum_{i=1}^{n} p_i(1-p_i)x_i.$$

Show that the asymptotic variance of $\hat{\beta}_1$ evaluated it at the MLE is

$$\widehat{\mathrm{Var}}(\hat{\beta}_1) = \frac{1}{n_{10}} + \frac{1}{n_{11}}.$$

Solution: Since $x_i^2 = x_i$ for $x_i \in \{0,1\}$ and only subjects with $x_i = 1$ contribute,

$$\mathcal{I}(\beta_1) = \sum_{i:\, x_i=1} p(1)\big(1-p(1)\big) = n_1\, p(1)\big(1-p(1)\big).$$

The asymptotic variance is $\mathrm{Var}(\hat{\beta}_1) = \mathcal{I}(\beta_1)^{-1}$. Evaluating at the MLE, $\hat{p}(1) = \frac{n_{11}}{n_1}$ and $1-\hat{p}(1) = \frac{n_{10}}{n_1}$, so

$$n_1\, \hat{p}(1)\big(1-\hat{p}(1)\big) = n_1 \cdot \frac{n_{11}}{n_1} \cdot \frac{n_{10}}{n_1} = \frac{n_{10}\, n_{11}}{n_1}.$$

Therefore

$$\widehat{\mathrm{Var}}(\hat{\beta}_1) = \frac{n_1}{n_{10}\, n_{11}} = \frac{n_{10}+n_{11}}{n_{10}\, n_{11}} = \boxed{\frac{1}{n_{10}} + \frac{1}{n_{11}}.}$$

(F) Consider the hypothesis test:

$$H_0 : \beta_1 = 0, \quad H_A : \beta_1 \neq 0.$$

Show that the Wald test statistic is

$$W = \hat{\beta}_1 \times \mathcal{I}(\hat{\beta}_1) \times \hat{\beta}_1 = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)\left(\log\frac{\hat{p}_i}{1-\hat{p}_i}\right)^2,$$

where $\hat{p}_i = p_i(\hat{\beta}_1)$. What is the null distribution of W?

Solution: The general Wald statistic for a scalar parameter is $W = \hat{\beta}_1 \cdot \mathcal{I}(\hat{\beta}_1) \cdot \hat{\beta}_1$. Since $\mathcal{I}(\hat{\beta}_1) = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)\, x_i$,

$$W = \hat{\beta}_1^2 \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)\, x_i = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)\, \hat{\beta}_1^2 x_i.$$

For $x_i = 0$: $\hat{p}_i = \frac{1}{2}$ so $\log\frac{\hat{p}_i}{1-\hat{p}_i} = 0$, and the summand vanishes. For $x_i = 1$: $\log\frac{\hat{p}_i}{1-\hat{p}_i} = \hat{\beta}_1 x_i = \hat{\beta}_1$. Hence in both cases $\hat{\beta}_1^2 x_i = \left(\log\frac{\hat{p}_i}{1-\hat{p}_i}\right)^2$, giving

$$\boxed{W = \sum_{i=1}^{n} \hat{p}_i(1-\hat{p}_i)\left(\log\frac{\hat{p}_i}{1-\hat{p}_i}\right)^2.}$$

By the asymptotic normality of the MLE, $\hat{\beta}_1\, [\mathcal{I}(\hat{\beta}_1)]^{1/2} \xrightarrow{d} N(0,1)$ under H_0, so $W \xrightarrow{d} \chi^2(1)$. We reject H_0 at level α when $W > \chi^2_{1,\alpha}$.