Homework 4

蒋翌坤 20307100013

5.4

(a)

Let kyphosis was present: $y = \pi(x) = 0$, kyphosis was absent: $y = \pi(x) = 1$, we want to fit the logistic regression model: $\text{logit}(\pi(x)) = \alpha + \beta x$, using software, we get the result:

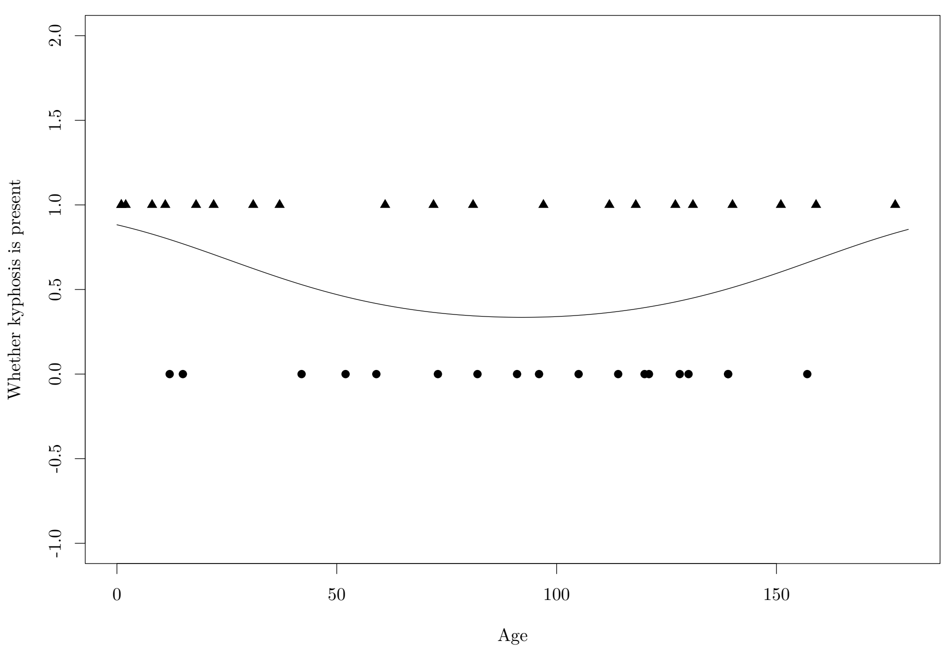
$\hat{\alpha} = 0.746, \hat{\beta} = -0.007, \text{SE}(\hat{\alpha}) = 0.628, \text{SE}(\hat{\beta}) = 0.006$

p-value for β is 0.26, so we can't reject the null hypothesis that $\beta = 0$, which means that age does not have a significant effect.

(b)

Using software to fit $\text{logit}(\pi(x)) = \alpha + \beta_1 x + \beta_2 x^2$, we get the result: $\text{logit}(\pi(x)) = 2.02 - 0.059x + 3.18 \times 10^{-4} x^2$. The p-value for β_2 is 0.0522, so there is only a faint significance of the squared age term.

Data plot and the fit is shown in the following Figure. We can see that when age is very small or large, the probability of kyphosis is absent is very high, while when age is around middle, the probability of kyphosis is present is very high.



Data plot and the fit for 5.4(b)

5.8

The estimated model is $\text{logit}(\pi_i) = \alpha - 1.320x_{1i} + 0.622x_{2i} + 0.501x_{3i} - 0.460x_{4i}$. p-value of all the parameters are all less than 0.05, so all parameters are significant. For age and marital status, they have a negetive effect on the use of oral contraceptives; For race and education, they have a positive effect on the use of oral contraceptives.

The confidence interval for conditional odds ratio between contraceptive use and education is $e^{0.501\pm1.96\times0.077} = (1.419, 1.919)$. Since the odds ratio is greater than 1, it shows the positive effect of education on the use of oral contraceptives.

5.32

(a)

$\text{logit}(\pi) = \log\frac{\pi}{1-\pi} = \alpha + \beta\log d \Rightarrow \frac{\pi}{1-\pi} = e^{\alpha+\beta\log d} = e^{\alpha}d^{\beta}$

odds for the first draft pick ($d = 1$): $\frac{\pi}{1-\pi} = e^{\alpha}1^{\beta} = e^{\alpha}$

(b)

The odds of all star is $e^{\alpha}d^{\beta}$, it decreases as β decreases; when it is the first draft, the odds are e^{α}, when α is larger, the probability of being an all star is larger. The basketball's $\hat{\beta}$ is smaller than baseball's $\hat{\beta}$, its $\hat{\alpha}$ is larger than baseball's $\hat{\alpha}$, so in basketball a first draft pick is more crucial and picks with high d are relatively less likely to be all stars.

5.36

We know $\pi(x) = \frac{1}{1+e^{-(\alpha+\beta x)}}$, when $x = 0$, $\pi(x) = \frac{1}{1+e^{-\alpha}}$, when $x = 1$, $\pi(x) = \frac{1}{1+e^{-(\alpha+\beta)}}$, so the likelyhood function is:

$$L = \Big(\frac{1}{1+e^{-\alpha}}\Big)^{y_0}\Big(1-\frac{1}{1+e^{-\alpha}}\Big)^{n_0-y_0}\Big(\frac{1}{1+e^{-(\alpha+\beta)}}\Big)^{y_1}\Big(1-\frac{1}{1+e^{-(\alpha+\beta)}}\Big)^{n_1-y_1}$$

The log likelyhood function is:

$$\begin{aligned} l &= y_0\log\Big(\frac{1}{1+e^{-\alpha}}\Big) + (n_0-y_0)\log\Big(\frac{1}{1+e^{\alpha}}\Big) + y_1\log\Big(\frac{1}{1+e^{-(\alpha+\beta)}}\Big) + (n_1-y_1)\log\Big(\frac{1}{1+e^{\alpha+\beta}}\Big) \\ &= -\Big(y_0\log(1+e^{-\alpha}) + (n_0-y_0)\log(1+e^{\alpha}) + y_1\log(1+e^{-(\alpha+\beta)}) + (n_1-y_1)\log(1+e^{\alpha+\beta})\Big) \end{aligned}$$

We have the likelihood equations:

$$\begin{cases} \dfrac{\partial l}{\partial\alpha} = -\Big(-\dfrac{y_0e^{-\alpha}}{1+e^{-\alpha}} + \dfrac{(n_0-y_0)e^{\alpha}}{1+e^{\alpha}} - \dfrac{y_1e^{-(\alpha+\beta)}}{1+e^{-(\alpha+\beta)}} + \dfrac{(n_1-y_1)e^{\alpha+\beta}}{1+e^{\alpha+\beta}}\Big) = 0 \\ \dfrac{\partial l}{\partial\beta} = -\Big(-\dfrac{y_1e^{-(\alpha+\beta)}}{1+e^{-(\alpha+\beta)}} + \dfrac{(n_1-y_1)e^{\alpha+\beta}}{1+e^{\alpha+\beta}}\Big) = 0 \end{cases}$$

Solve the equations, we get $\begin{cases} \hat{\alpha} = \log\frac{y_0}{n_0-y_0} \\ \hat{\beta} = \log\frac{y_1}{n_1-y_1} - \log\frac{y_0}{n_0-y_0} \end{cases}$.