Solution for CDA2015

- **Solution to 5.4**. (a). Let $\pi(x)$ be the probability that kyphosis is present for patient with age in month x. Fit the model $\text{logit}\pi(x) = \alpha + \beta x$. The output is

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
α	-0.572693106	0.60239457	-0.9506943	0.3417596
β	0.004295787	0.00584936	0.7344029	0.4627032

$\text{logit}[\pi(x)] = 0.5727 - 0.0043x$, $p-$value=0.4627. Not reject H_0.

(b). Fit the model $\text{logit}[\pi(x)] = \alpha + \beta_1 x + \beta_2 x^2$. The output is

	Estimate	Std. Error	z value	$Pr(>\|z\|)$
α	-2.0462546726	0.9943478097	-2.057886	0.03960106
β_1	0.0600397937	0.0267808473	2.241893	0.02496829
β_2	-0.0003279439	0.0001564095	-2.096700	0.03602012

So

$$\hat{\pi}(x) \approx \frac{\exp(-2.0462546726 + 0.0600397937x - 0.0003279439x^2)}{1 + \exp(-2.0462546726 + 0.0600397937x - 0.0003279439x^2)}.$$

For testing $H_0 : \beta_2 = 0$, Wald statistic $z = \hat{\beta}_2/SE(\hat{\beta}_2) \approx -2.096700$, and $p-$value =0.03602012. So reject H_0 and the item of squared age is significant.

$\text{logit}[\pi(x)] = 2.0163 - 0.0600x + 0.0003x^2$, $p-$value=0.036.

- **Solution to 5.8**. The estimated model is

$$\text{logit}(\hat{\pi}_i) = \alpha - 1.320x_{1i} + 0.622x_{2i} + 0.501x_{3i} - 0.460x_{4i},$$

where x_{1i}, x_{2i}, x_{3i} and x_{4i} are the dummy variables of object i for age, race, education and married status, respectively. The value of α is not listed in the table. It is obvious that all of absolute values of parameters are larger than their standard values respectively. So, all the factors have effect on the use of oral contraceptives. Age and martial status have negative effect (that means less young women or married women use oral contraceptives), while race and education have positive effect (more white women or women with high education use oral contraceptive).

The point estimation and 95% CI for the conditional odds ratio between contraceptives are $\hat{\theta} = \exp(0.501) = 1.6504$ and $e^{0.501 \pm 1.96 * 0.077} = (1.4192,\ 1.9192)$, respectively. So, conditional on other factors, education and use of oral contraceptives are not independent.

- **Solution to 5.32**. (a). $\frac{\pi}{1-\pi} = e^{\alpha + \beta \log d} = e^{\alpha} d^{\beta}$. $d = 1$ implies that $\frac{\pi}{1-\pi} = e^{\alpha}$.

(b). Since $\hat{\pi} = \frac{e^{\hat{\alpha}} d^{\hat{\beta}}}{1 + e^{\hat{\alpha}} d^{\hat{\beta}}}$, it follows that $\hat{\pi}$ increases as $\hat{\alpha}$ increasing; that $\hat{\pi}$ increases as $\hat{\beta}$ increasing; and that $\hat{\pi}$ decreases as d increasing.

- **Solution to 5.36**. The likelihood function is

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

So,

$$\log L(\alpha,\beta)=(y_0+y_1)\alpha+y_1\beta-n_0\log(1+e^{\alpha})-n_1\log(1+e^{\alpha+\beta}).$$

Hence, the ML estimators for α and β are

$$\hat{\alpha}=\operatorname{logit}(y_0/n_0),\quad \hat{\beta}=\operatorname{logit}(y_1/n_1)-\operatorname{logit}(y_0/n_0).$$

- **Solution to 6.6**. For the model $\operatorname{logit}(\pi(x))=\alpha$, we have $\hat{\alpha}=-1.0186$, $G^2=7.2777$ with $df=4$ and p-value$=0.122>0.05$. The residuals are below.

	Fitted	Reschi	Resdev	StReschi
Worse	3.183673	-1.5975945	-1.4278094	-1.4736330
Stationary	17.510204	-1.2976089	-1.2574693	-1.5440229
Slight improvement	15.387755	0.1813343	0.1820893	0.2170065
Moderate improvement	11.142857	1.3064649	1.3480756	1.5208343
Marked improvement	4.775510	1.1409986	1.1875922	1.2461931

Since all $|\text{residuals}|<2$, the model fits very well and no lack of fit.

(Note: some students say there exists increasing trend. This is not true since data points are small and not strictly increasing.)

- **Solution to 6.22**. (a). Note that $G^2(M_j|M_k)+G^2(M_k|M_s)=G^2(M_j|M_s)$ for $j<k<s$ and $G^2(M_k|M_s)\geq 0$. So $G^2(M_j|M_k)\leq G^2(M_j|M_s)$.

(b). Note that $G^2(M_j|M_k)\leq G^2(M_j|M_s)\leq G^2(M_1|M_s)$ and $G^2(M_1|M_s)\xrightarrow{d}\chi_v^2$. So,

$$P[G^2(M_j|M_k)>\chi_v^2(\alpha)]\leq P[G^2(M_1|M_s)>\chi_v^2(\alpha)]\to\alpha.$$

So $P[G^2(M_j|M_k)>\chi_v^2(\alpha)]\leq\alpha$ for all $k>j$.

(c).

$$\begin{aligned}&P[G^2(M_j|M_k)\leq\chi_v^2(\alpha)\text{ for all }j<k]\\&=1-P[G^2(M_j|M_k)>\chi_v^2(\alpha)\text{ for some }j<k]\\&\geq 1-P[G^2(M_1|M_s)>\chi_v^2(\alpha)]\geq 1-\alpha.\end{aligned}$$

- **Solution to 6.24**. (a). The model is

$$\operatorname{logit}(\pi_{ik})=\alpha+\beta x_i+\beta_k^Z,\quad \pi_{ik}=\frac{\exp(\alpha+\beta x_i+\beta_k^Z)}{1+\exp(\alpha+\beta x_i+\beta_k^Z)},\quad x_i=\begin{cases}1 & , \quad i=1\\ 0 & , \quad i=2\end{cases}$$

So, the likelihood function is

$$\mathcal{L} = \prod_{i=1}^{2}\prod_{k=1}^{K} \pi_{ik}^{n_{i1k}}(1-\pi_{ik})^{n_{i2k}} = \prod_{i=1}^{2}\prod_{k=1}^{K} \frac{[\exp(\alpha+\beta x_i+\beta_k^Z)]^{n_{i1k}}}{[1+\exp(\alpha+\beta x_i+\beta_k^Z)]^{n_{i+k}}}$$

and

$$\log\mathcal{L} = \sum_{i=1}^{2}\sum_{k=1}^{K} n_{i1k}(\alpha+\beta x_i+\beta_k^Z) - \sum_{i=1}^{2}\sum_{k=1}^{K} n_{i+k}\log[1+\exp(\alpha+\beta x_i+\beta_k^Z)].$$

So α's sufficient statistic is n_{+1+}, β's sufficient statistic is n_{11+}, and β_k^Z's sufficient statistic is n_{+1k}.

(b) Skip it, since we do not study uniformly most powerful unbiased tests and sufficient statistics.

- **Solution to 6.24**. (a). For Probit model

$$\pi_i = \Phi\left(\sum_j \beta_j x_{ij}\right),$$

the log likelihood is

$$\begin{aligned} L_i &= y_i\log\pi_i + (1-y_i)\log(1-\pi_i) \\ &= y_i\log\Phi\left(\sum_j \beta_j x_{ij}\right) + (1-y_i)\log\left[1-\Phi\left(\sum_j \beta_j x_{ij}\right)\right]. \end{aligned}$$

So,

$$L = \sum_i L_i = \sum_i y_i\log\Phi\left(\sum_j \beta_j x_{ij}\right) + \sum_i (1-y_i)\log\left[1-\Phi\left(\sum_j \beta_j x_{ij}\right)\right].$$

(b). Since

$$\frac{\partial L_i}{\partial\beta_j} = \frac{(y_i-\pi_i)x_{ij}}{\pi_i(1-\pi_i)}\phi\left(\sum_j \beta_j x_{ij}\right),$$

the estimation equation is

$$\frac{\partial L}{\partial\beta_j} = \sum_i \frac{\partial L_i}{\partial\beta_j} = \sum_i \frac{(y_i-\pi_i)x_{ij}}{\pi_i(1-\pi_i)}\phi\left(\sum_j \beta_j x_{ij}\right) = 0, j=0,\dots,p.$$

Hence the estimation equation can be written as

$$\sum_i (y_i-\hat{\pi}_i)z_i x_{ij} = 0, j=0,\dots,p,$$

where

$$z_i = \frac{\phi\left(\sum_j \hat{\beta}_j x_{ij}\right)}{\hat{\pi}_i(1-\hat{\pi}_i)}.$$

For Logistic model, it follows that

$$\hat{\eta}_i = \log \frac{\hat{\pi}_i}{1 - \hat{\pi}_i} = \sum_j \hat{\beta}_j x_{ij}.$$

Since $\phi(x) = \frac{e^x}{1+e^x} \frac{1}{1+e^x}$, then

$$\phi(\hat{\eta}_i) = \frac{e^{\hat{\eta}_i}}{1 + e^{\hat{\eta}_i}} \frac{1}{1 + e^{\hat{\eta}_i}} = \hat{\pi}_i (1 - \hat{\pi}_i),$$

which implies $z_i = 1$.