Review 2: Chapters 4, 5 & 6

1. Logistic regression model & interpretation:

   For data $Y \sim Bin(n_i, \pi(x))$, logistic regression model

$$\text{logit}\{\pi(x)\} = \log\left\{\frac{\pi(x)}{1-\pi(x)}\right\} = \alpha + \beta x.$$

$$\begin{aligned}
\alpha &= \log\left\{\frac{\pi(0)}{1-\pi(0)}\right\} : \text{log odds of success at } x = 0 \\
\pi(0) &= \frac{e^\alpha}{1+e^\alpha}. \\
\beta &= \log\left\{\frac{\pi(x+1)/\{1-\pi(x+1)\}}{\pi(x)/\{1-\pi(x)\}}\right\} \\
&\quad \text{log odds-ratio of success with 1 unit increase of } x \\
e^\beta &= \frac{\pi(x+1)/\{1-\pi(x+1)\}}{\pi(x)/\{1-\pi(x)\}} \\
&\quad \text{odds-ratio of success with 1 unit increase of } x
\end{aligned}$$

$$\pi'(x) = \beta\pi(x)\{1-\pi(x)\}.$$

2. Confidence interval of $\pi(x_0)$:

$$\pi(x_0) = \frac{e^{\eta(x_0)}}{1+e^{\eta(x_0)}}.$$

   $(1-\alpha)$ CI for $\eta(x_0) = \alpha + \beta x_0$: $[\widehat{\eta}_1, \widehat{\eta}_2]$, then $(1-\alpha)$ CI for $\pi(x_0)$:

$$\left[\frac{e^{\widehat{\eta}_1}}{1+e^{\widehat{\eta}_1}}, \frac{e^{\widehat{\eta}_2}}{1+e^{\widehat{\eta}_2}}\right].$$

3. Logistic models with multiple x's, interpretation of their coefficients, CI for $\pi(x_o)$, etc.

4. Inference for $2\times2\times K$ tables, logistic regression model (K is small to moderate), conditional logistic regression model (good for large K) & CMH test (how to construct, good for larger K).

5. Model checking for logistic regression model: Pearson χ^2 & Deviance tests (when is it valid?), Hosmer-Lemeshow test (for continuous x's), Pearson residuals and standardized Pearson residuals

6. ROC curve, AUC of ROC (c-index)

7. Sparse data, exact inference:

   (a) $2 \times 2 \times K$ tables, exact CMH test (SAS program?)

   (b) Ordinal or continuous x, exact Cochran-Armitage trend test (SAS program?)

8. Sample size calculation:
$$n_1 = n_2 = \frac{(z_{\alpha/2} + z_\beta)^2[\pi_1(1-\pi_1) + \pi_2(1-\pi_2)]}{(\pi_1 - \pi_2)^2}.$$

9. Baseline category logit models (for nominal Y):
$$\log\left\{\frac{\pi_j(x_i)}{\pi_J(x_i)}\right\} = \alpha_j + \beta_j x_i, \quad j = 1, 2, ..., J-1$$

   Cell probabilities:
$$\begin{aligned}
\pi_j(x) &= \frac{e^{\alpha_j + \beta_j x}}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}} \quad j = 1, 2, ..., J-1, \\
\pi_J(x) &= \frac{1}{1 + \sum_{k=1}^{J-1} e^{\alpha_k + \beta_k x}}
\end{aligned}$$

   GOF: Deviance, Pearson χ^2, $df = (J-1) \times [(I-1) - \#\text{ of } x\text{'s}]$.

10. Cumulative logit models (for ordinal Y):
$$\tau_j(x) = P[Y \le j|x] = \pi_1(x) + \pi_2(x) + \cdots + \pi_j(x), \quad j = 1, 2, ..., J-1,$$
$$\log\left\{\frac{\tau_j(x)}{1 - \tau_j(x)}\right\} = \alpha_j + \beta x, \quad j = 1, 2, ..., J-1$$

   Cell probabilities:
$$\tau_j(x) = \frac{e^{\alpha_j + \beta x}}{1 + e^{\alpha_j + \beta x}}, \quad j = 1, 2..., J-1$$
$$\Rightarrow$$
$$\begin{aligned}
&\pi_1(x) = \tau_1(x) \\
&\pi_2(x) = \tau_2(x) - \tau_1(x) \\
&... \\
&\pi_j(x) = \tau_j(x) - \tau_{j-1}(x) \\
&... \\
&\pi_{J-1}(x) = \tau_{J-1}(x) - \tau_{J-2}(x) \\
&\pi_J(x) = 1 - \tau_{J-1}(x)
\end{aligned}$$

GOF:

(a) Score test, H_a :

$$\log \left\{ \frac{\tau_j(x)}{1 - \tau_j(x)} \right\} = \alpha_j + \beta_j x, \quad j = 1, 2, ..., J-1$$

$df = (J-2) \times dim(x)$

(b) Deviance or Pearson χ^2 test (when is it valid): $df = (I-1)(J-1) - dim(x)$