G. Parameters are log of relative risk for individuals with covariate 1 unit apart, identical otherwise.

i. Likelihood is

\[ L(\alpha) = \left( \frac{X_+}{X_0, \ldots, X_{K-1}} \right) \times \prod_{k=0}^{K-1} \exp([\omega_k + \alpha_k]X_k) \left[ \sum_{m=0}^{K-1} \exp(\omega_m + \alpha_m) \right]^{-X_+}. \]

ii. Log likelihood is

\[ l(\alpha) = \log \left( \frac{X_+}{X_0, \ldots, X_{K-1}} \right) + \sum_{k=0}^{K-1} [\omega_k + \alpha_k]X_k - X_+ \log \left[ \sum_{m=0}^{K-1} \exp(\omega_m + \alpha_m) \right]. \]

iii. \( l^k(\alpha) = X_k - X_+ \frac{\exp(\omega_k + \alpha_k)}{\sum_{m=0}^{K-1} \exp(\omega_m + \alpha_m)}. \)

iv. Score is \( l^k(\theta) = X_k - X_+ \exp(\omega_k) / \left[ \sum_{m=0}^{K-1} \exp(\omega_m) \right]. \)

1. Parameter estimates are logs of relative risk

2. Testing done via
   a. standard errors, which come from Delta method (Wald test)
      i. Also gives CI

     \[ \text{B&G D2: 4.3c–d} \]

   b. likelihood ratio

3. Complications:
   a. Do iterations bounce back and forth without converging?
   b. Sometimes best fits for parameters are \( \pm \infty \)
4. Does model fit well?
   a. Predicted mean values for each of the groups ought to be about right
   b. Hence $\sum_k (X_k - e_k)^2 / e_k$ ought to be approximately $\chi^2$
      i. DF is number of groups - number of parameters
   c. Alternatively, use likelihood ratio
      i. Write down probability for data
      ii. Express as function of unknown parameters
         • Function $L$ is called likelihood.
      iii. Parameter value that maximizes $L$ is called the maximum likelihood estimate
      iv. $H_0$ is plausible if $L$ is not much higher somewhere else.
      v. Hence test hypothesis by comparing maximized value to value at null
         • compare with ratio to get likelihood ratio test
         • usually take log: $l = \log(L)$.
         • $2 \times$ difference in $l$ generally approximately $\sim \chi_k^2$ for $k$ the
difference in number of unknown parameters.
H. Testing parameter values is done via

1. standard errors, which come from Delta method (Wald test)
   a. Also gives CI

2. likelihood ratio
   a. Write down probability for data
   b. Express as function of unknown parameters
      i. Function $L$ is called likelihood.
   c. Parameter value that maximizes $L$ is called the maximum likelihood estimate
   d. $H_0$ is plausible if $L$ is not much higher somewhere else.
   e. Hence test hypothesis by comparing maximized value to value at null
      i. compare with ratio to get likelihood ratio test
      ii. usually take log: $l = \log(L)$.
      iii. $2 \times$ difference in $l$ generally approximately $\sim \chi^2_k$ for $k$ the difference in number of unknown parameters.
   f. $-2 \times l$ is called deviance
i. after subtracting off $-2 \times \log$ likelihood for model with a separate rate for each line in data set

ii. Bigger model is called \textit{saturated model}.

I. Does model fit well?

1. Predicted mean values for each of the groups ought to be about right

2. Hence $\sum_j (n_{jk} - E_{\hat{\beta}}[n_j])^2/Var_{\hat{\beta}}[n_j]$ ought to be approximately $\chi^2$

   a. For Poisson regression, $E_{\hat{\beta}}[n_j] = Var_{\hat{\beta}}[n_j] = \exp(z_k\hat{\beta})Q_j$

   b. DF is number of groups - number of parameters

3. Alternatively, use likelihood ratio

   a. Embed in bigger model where every observation gets its own parameter value

      \textbf{A1: 6-6.1}

J. Two Dimensions: $\lambda_{kj} = \exp(\mu + \alpha_k + \gamma_j + \sigma_{kj})$

1. Clearly saturated: Could get any set of rates we want by setting $\sigma_{kj}$ equal to the associated rate.

2. Overparameterized: fix by

   a. setting $\sigma_{kj} = 0$ if $j = 0$ or $k = 0$. 
b. Still over parameterized, since increasing $\alpha$'s by a constant, and decreasing the $\gamma$'s by the same amount, leaves model fits unchanged.

3. Odds ratio $\theta_{k,j} = \frac{\exp(\mu + \alpha_k + \gamma_j + \sigma_{k,j}) \exp(\mu + \alpha_0 + \gamma_0 + \sigma_{00})}{\exp(\mu + \alpha_k + \gamma_0 + \sigma_{k0}) \exp(\mu + \alpha_0 + \gamma_j + \sigma_{00})} = \exp(\sigma_{k,j})$.

4. Model with no interactions: $\sigma_{k,j} = 0$ for all $k, j$
   a. Fix overparameterization by requiring $\exp(\sum_j \gamma_j) = 1, \mu = 0$
   b. $\frac{dl(\alpha, \gamma)}{d\alpha_k} = X_k+ - \sum_j \exp(\alpha_k + \gamma_j) = X_k+ - \exp(\alpha_k)$
      i. Hence $\hat{\alpha}_k = \log(X_k+)$
   c. $\frac{dl(\alpha, \gamma)}{d\gamma_j} = X_{+j} - \sum_k \exp(\alpha_k + \gamma_j) = X_{+j} - \exp(\gamma_j) \sum_k \exp(\alpha_k)$
      i. Hence $\hat{\gamma}_j = \log(X_{+j} / X_{++})$
   d. Pearson goodness of fit statistic is test of $H_0: \sigma_{k,j} = 0 \forall j, k$, with $\alpha, \gamma$ arbitrary.
      i. Only if data are arranged so that each cell in the full table is represented by one line in the data file.

5. Model with interactions:
   a. Easiest constraint: $\mu = 0$, $\alpha_j = \gamma_k = 0$ for all $k, j$
b. $\hat{\sigma}_{k,j} = \log(X_{k,j})$

c. Under our earlier constraint that $\sigma_{k,j} = 0$ if $j = 0$, get
   $$\hat{\sigma}_{k,j} = \log(X_{k,j}) - \log(X_{k0})$$

d. Under additional constraint that $\sigma_{k,j} = 0$ if $k = 0$, get
   $$\hat{\sigma}_{k,j} = \log(X_{k,j}) - \log(X_{k0}) - \log(X_{0j}) + \log(X_{00})$$:
   Earlier log odds ratio.  

K. Three Dimensions: $\lambda_{k,j}^i = \exp(\sigma_{k,j}^i)$

1. Again, $\hat{\sigma}_{k,j}^i = \log(X_{k,j}^i)$

2. Model forming null hypothesis for cmh test: $\lambda_{k,j}^i = \exp(\tau_{k,j}^i + \zeta_{j}^i + \xi_{i,k}^i)$

   a. No closed form for MLE

VI. Regression models for probabilities instead of rates

   B&D2: 4.7

A. Proportional Mortality

1. How do risks of two (mutually exclusive) events compare?

   a. Common offset $\omega_k$

   b. Assume $x_{k,j} \sim \mathcal{P}(\lambda_{k,j})$, $\lambda_{k,j} = \exp(\omega_k + z_{k,j}\beta)$.

   c. Then $X_{k,1|X_{k,+}} = x_{k,+} \sim \text{Bin}(\pi_{k1}, x_{k,+})$ for
Lecture 6

\[ \pi_{k1} = \frac{\lambda_{k1}}{\lambda_{k1} + \lambda_{k2}} \]

d. \[ \pi_{k1} = \frac{\exp(\omega_k + z_{k1}\beta)}{\exp(\omega_k + z_{k1}\beta) + \exp(\omega_k + z_{k2}\beta)} = \exp((z_{k1} - z_{k2})\beta) / [\exp((z_{k1} - z_{k2})\beta) + 1] \]

A: 3.2.3

e. \[ \text{logit}(\pi_{k1}) = \Delta \beta \]

f. Method is called \textit{logistic regression}

g. Standard errors come from delta method

B. Parameter estimates are logs of odds for individuals with covariate 1 unit apart, identical otherwise.

C. Complications:

1. Do iterations bounce back and forth without converging?
2. Sometimes best fits for parameters are \( \pm \infty \)
3. Tests can mislead when some groups have small expected value

D. Logistic regression for \( K \times 2 \) tables:

1. \[ x_{k1}\mid X_{k+} = x_{k+} \sim \text{Bin}(x_{k+}, 1/(1 + \exp(-\mu - \alpha_k))) \]
2. For \( 2 \times 2 \) table analysis, cohort study (exposed and unexposed group sizes fixed)
a. Recall notation: $x_{kj}^i$ is number of \(\begin{cases} \text{cases} & \text{if } j = 1 \\ \text{controls} & \text{if } j = 0 \end{cases}\) at exposure level \(\begin{cases} \text{exposed} & \text{if } k = 1 \\ \text{none} & \text{if } k = 0 \end{cases}\) in strata \(i\) (if needed)

b. Expression as binomials

i. Number of cases among unexposed \((k=0)\) and exposed \((k=1)\) is 
\[ X_{k1} | X_{k+} = x_{k+} \sim \text{Bin}(\pi_k, x_{k+}) \]

c. Write as regression model

i. \(\text{logit}(\pi_0) = \mu\)

ii. \(\text{logit}(\pi_1) = \log(\pi_1/(1-\pi_1)) = \log(\pi_0/(1-\pi_0)) + \log(\theta) = \mu + \alpha_1 \text{ for } \alpha_1 = \log(\theta)\).

d. Recall conditioning on \(X_{1+}\) removes effect of \(\mu\)

3. We have too many parameters

a. Can decrease \(\mu\) and increase each other \(\alpha_1\) and get same probabilities

b. Three typical solutions:

i. Set \(\mu = 0\): Results in separate log odds fits for each row.

ii. Set \(\sum_{k=0}^{K-1} \alpha_k = 0\): Makes \(\mu\) an “average” log odds, and rest are log odds ratios in comparison to average.

iii. Set \(\alpha_{k'} = 0\) for some \(k' \in \{0, \ldots, K - 1\}\).

- Makes group \(k'\) the reference group
• $\mu$ represents log odds for reference group
• $\alpha_k$ is the log odds for group $k$ with respect to group $k'$. 
• Typically choose $k'$ as 1 or $K$.

c. Unlike contingency table approach, this approach is not conditional on number with disease. [Mark D sas] [Mark D R]

A: 4.3.4

4. We can use this approach for stratified $K \times 2$ tables

a. to estimate common odds ratios
b. to test whether odds ratio is really constant.
   i. non-constant odds ratio is equivalent to interactions between effect and stratification variable

c. Unlike Mantel–Haenzel approach, this approach is not conditional on disease numbers in each table.

5. Approach can be extended to scored categories.

a. Add in score as a covariate [Mark E sas] [Mark E R]

A: 4.1

6. Continuous covariates may also be used

a. As with simpler regression models, one should consider the proper scale for continuous covariates
b. Consider adding polynomial terms

c. Estimates could be seriously impacted by other variables in model

  i. mild effect of collinearity
  
  ii. Impact can be minimized by subtracting out mean.

St: 11.2

7. You can use another function instead of logit

   a. Must still map $\mathcal{R}$ into $[0, 1]$  
   
   b. Logit has some mathematical properties we will discuss later  
   
   c. Normal CDF is sometimes used  

      i. Called the probit

      ii. Results from discretizing standard multiple regression.

      • Suppose $Y_j = x_j \beta + \sigma \epsilon_j$, $\epsilon_j \sim N(0, 1)$
      • $Z_j = \begin{cases} 1 & \text{if } Y_j > c \\ 0 & \text{otherwise} \end{cases}$
      • $P[Z_j = 1] = P[Y_j > c] = P[\epsilon_j > (c - x_j \beta)/\sigma] = 1 - \Phi((c - x_j \beta)/\sigma) = \Phi((x_j \beta - c)/\sigma)$

      • After rescaling, probit and logit are very close. See Figure 8.
Fig. 8: Comparison of Probit and Logistic Link Functions

\[
\begin{align*}
\text{Probit } & \Phi(x) \\
\text{Logistic } & \frac{\exp(1.702x)}{1 + \exp(1.702x)}
\end{align*}
\]