- Heuristic explanation: Rates for $(1, 0, 0)$ and $(0, 1, 1)$ are the same, and so can't tell difference between them.
- Problem is called colinearity

B&D2: 4.6

7. Model contains log of time at risk as an *offset*
   a. Fit component is added to every log rate
   b. If you know something that rates might be proportional to, log of this could be added to the offset as well
      i. For ex, rate in unexposed population by age
8. Parameters are log of relative risk for individuals with covariate 1 unit apart, identical otherwise.
9. Testing parameter values is done via
   a. standard errors, which come from Delta method (Wald test)
      i. Also gives CI

B&D1: 6.4

   b. 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 $\sim \chi^2_k$ for k the difference in number of unknown parameters.

vi. $-2 \times l$ is called *deviance*

- after subtracting off $-2\times$ log likelihood for model with a separate rate for each line in data set
- Bigger model is called *saturated model*.

10. Does model fit well?

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

b. Hence $\sum_j (O_{jk} - \mathrm{E}_{\hat{\boldsymbol{\beta}}}[O_j])^2 / \mathrm{Var}_{\hat{\boldsymbol{\beta}}}[O_j]$ ought to be approximately χ^2

i. For Poisson regression, $\mathrm{E}_{\hat{\boldsymbol{\beta}}}[O_j] = \mathrm{Var}_{\hat{\boldsymbol{\beta}}}[O_j] =$

$\exp(\boldsymbol{x}_k\hat{\boldsymbol{\beta}})Q_j$

ii. DF is number of groups - number of parameters

c. Alternatively, use likelihood ratio

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

F. Regression models for probabilities instead of rates

B&D2: 4.7

1. Proportional Mortality

a. What if we don't have person–years at risk?

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

i. Assume ${O_k}^1 \sim \mathcal{P}(\lambda_k)$, ${O_k}^2 \sim \mathcal{P}(\nu_k)$

ii. Then ${O_k}^1|{O_k}^+ \sim$ Bin$(\pi_k, {O_k}^+)$ for $\pi_k = \lambda_k/(\lambda_k + \nu_k)$

iii. $\pi_k = \exp(\boldsymbol{x}_k\boldsymbol{\beta})/[\exp(\boldsymbol{x}_k\boldsymbol{\beta}) + \exp(\boldsymbol{x}_{jk}\boldsymbol{\delta})] = \exp(\boldsymbol{x}_{jk}(\boldsymbol{\beta} - \boldsymbol{\delta}))/[\exp(\boldsymbol{x}_{jk}(\boldsymbol{\beta} - \boldsymbol{\delta})) + 1]$

iv. If second type of event does not depend on exposure, then $\boldsymbol{\delta} = \mathbf{0}$, and $\boldsymbol{\beta} - \boldsymbol{\delta} = \boldsymbol{\beta}$

Se: 8 pp. 243–262

v. $\operatorname{logit}(\pi_{jk}) = \boldsymbol{x}_k\boldsymbol{\beta}$

vi. Method is called *logistic regression*

vii. Standard errors come from delta method

2. Fitting the model:

a. Start with a guess of best values for $\boldsymbol{\beta}$

i. Call them $\boldsymbol{\beta}^0$

ii. Almost any value (like $\mathbf{0}$) will do.

b. If z close to y then expand $\exp(z)/(1+\exp(z)$ as Taylor series

c. Then

$$O_{1j} = O_{+j}\pi_{1j} + \sqrt{O_{+j}\pi_j(1-\pi_j)}\epsilon_j$$
$$\approx O_{+j}\pi_j^0(1+(1-\pi_j^0)\boldsymbol{x}_j(\boldsymbol{\beta}-\boldsymbol{\beta}^0))+$$
$$\sqrt{O_{+j}\pi_j^0(1-\pi_j^0)}\epsilon_j$$

d. Hence

$$\frac{O_{1j}-O_{+j}\pi_j^0}{\sqrt{O_{+j}\pi_j^0(1-\pi_j^0)}} \approx \sqrt{O_{+j}\pi_j^0(1-\pi_j^0)}\boldsymbol{x}_j(\boldsymbol{\beta}-\boldsymbol{\beta}^0)+\epsilon_j$$

i. $\pi_j^0 = 1/(1+\exp(-\boldsymbol{x}_j\boldsymbol{\beta}^0))$

ii. $\epsilon_j \sim \mathcal{N}(0,1)$

iii. Now this looks like a regular regression problem

e. Use multiple regression to update guess

i. Do multiple times

ii. Method is called $iteratively\ reweighted\ least\ squares$.

3. Parameter estimates are logs of odds for individuals with covariate 1 unit apart, identical otherwise.
4. Complications:
   a. Do iterations bounce back and forth without converging?
   b. Sometimes best fits for parameters are $\pm\infty$
   c. Tests can mislead when some groups have small expected value
5. Problematic Examples
   a. Cohort Study with Common Disease
      i. Poisson methods fail
         - Counts of cases large enough to be influenced by finiteness of population are not rare enough
   b. Studies with rates that vary quickly with age,
      i. changing rate is accounted for by using age interval as class variable and modeling relation between class levels.
      ii. 960-542 provides more powerful and natural ways to model dependence of rate on time

Se: 7 pp. 214–220

6. Logistic regression for $K \times 2$ tables:
   a. $O_{k1}|O_{k+} \sim \text{Bin}(O_{k+}, 1/(1 + \exp(-\beta_0 - \beta_k))$

b. For 2×2 table analysis, cohort study (exposed and unexposed group sizes fixed)

i. Recall notation: ${O_{kj}}^i$ is number of $\begin{cases} cases & \text{if } j = 1 \\ controls & \text{if } j = 0 \end{cases}$ at exposure level $\begin{cases} exposed & \text{if } k = 1 \\ none & \text{if } k = 0 \end{cases}$ in strata i (if needed)

ii. Expression as binomials

- Number of cases among unexposed is $O_{01} \sim \text{Bin}(\pi_0, O_{0+})$
- Number of cases among exposed is $O_{11} \sim \text{Bin}(\pi_1, O_{1+})$

iii. Write as regression model

- $\text{logit}(\pi_0) = \beta_0$
- $\text{logit}(\pi_1) = \log(\pi_1/(1-\pi_1)) = \log(\pi_0/(1-\pi_0)) + \log(\psi) = \beta_0 + \beta_1$ for $\beta_1 = \log(\psi)$.

iv. Recall we conditioned on O_{1+} to remove effect of β_0

c. We have too many parameters

i. Can decrease β_0 and increase each other β_k and get same probabilities

ii. Three typical solutions:

- Set $\beta_0 = 0$: Results in separate log odds fits for each row.
- Set $\sum_{k=1}^{K} \beta_k = 0$: Makes β_0 an "average" log odds, and rest are log odds ratios in comparison to average.

- Set $\beta_{k'} = 0$ for some $k' \in \{1, \ldots, K\}$.
  - ▷ Makes group k' the reference group
  - ▷ β_0 represents log odds for reference group
  - ▷ β_k is the log odds for group k with respect to group k' .
  - ▷ Typically choose k' as 1 or K .

iii. Unlike contingency table approach, this approach is not conditional on number with disease.

Se: 7 pp. 220–229

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

i. to estimate common odds ratios

ii. to test whether odds ratio is really constant.

- non-constant odds ratio is equivalent to interactions between effect and stratification variable

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

e. Approach can be extended to scored categories.

i. Add in score as a covariate