- Modeling and Analysis of a 1:M MCCD with n matched sets ← n blocks
  - For individual i in the jth matched set $W=w_j$, $j=1, \ldots, n$, observe the value of risk factors x_{ij} ← from jth case; # of cases
  - Denote $i=0 \Rightarrow$ case and $i=1,\ldots, M \Rightarrow$ control

case → x_{0j}; controls → $x_{1j}, \ldots, x_{Mj}$

binary resp. → $z_{0j}, z_{1j}, \ldots, z_{Mj}$, with $z_{0j}=1$, $z_{1j}=0$, …, $z_{Mj}=0$

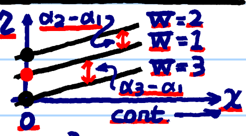
  - Assume the main-effect model of X and W (i.e., no interactions):

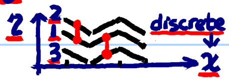
like an intercept in jth block; main effects of W; main effects of X

↳ treated as a qualitative variable with n levels (why?)

$$\text{logit}[\,p(\boldsymbol{w}_j, \boldsymbol{x}_{ij})\,] \equiv \text{logit}(p_{ij}) = \eta_{ij} = \alpha_j + \boldsymbol{x}_{ij}{}^T\beta$$

$P(Z_{ij}=1 \mid W=w_j, X=x_{ij})$ under logit link; effects of no interest; can be changed to $f_k(x)$'s; effects of main interest





(Δ) α_j's are not the true block effects β_{0j}'s in MCCD (why?) Check (*) in LNp 3-24, $\alpha_j = \log(e_j) + \beta_{0j}$

  α_j: effect of $W=w_j$ (jth match set)

$$p_{ij} = \frac{e^{\eta_{ij}}}{1+e^{\eta_{ij}}} \overset{*}{=} \frac{\exp(\alpha_j + x_{ij}^T\beta)}{1+\exp(\alpha_j + x_{ij}^T\beta)}, \quad 1-p_{ij} = \frac{1}{1+\exp(\cdots)}$$

  - Let $S_j = z_{0j}+z_{1j}+\ldots+z_{Mj}$ = the number of 1's (i.e., D) in the M+1 binary responses observed in the jth matched set
  $\Rightarrow S_j$: sufficient statistics of α_j

a block or a sub-population → $W=w_j$ ← Suppose random sampling from it or using Approach1 in LNp.3-42. Then S_j/K_j can be used to estimate the difference btwn β_{0j}'s

	D^c	D	
X^c	y_{j00}	y_{j01}	
X	y_{j10}	y_{j11}	
	$K_j - S_j$	S_j	K_j
In MCCD →	M	1	$M+1$

adjusted for in design

$\hat{\alpha}_j \leftarrow S_j/K_j = \frac{1}{M+1}$ all the time ⇒ cannot estimate

joint pmf of Z_{ij}'s, $i=0,\ldots,M$, $j=1,\ldots,n$,

$$= \prod_{i,j} p_{ij}^{z_{ij}}(1-p_{ij})^{1-z_{ij}} = \prod_{i,j} \exp[(\alpha_j + x_{ij}^T\beta)\times z_{ij}] \Big/ [1+\exp(\cdots)]^{n\times(M+1)}$$

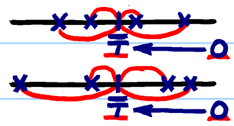
$$\boxed{S_j} \propto \exp\left[\textstyle\sum_j \alpha_j \left(\sum_i z_{ij}\right)\right] \times \exp\left[\left(\textstyle\sum_{i,j} z_{ij} x_{ij}\right)^T \beta\right]$$

- Regard α_j's as nuisance parameters, it is natural to make inference conditional on its sufficient statistics S_j's ← parameters of no interest; definition

can be applied on data collected from Approach1 in LNp.3-42 under the model in LNp.3-44

Consider the example.
- $T_1, \cdots, T_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$
  ↳ contain information about μ & σ^2
- $\bar{T}$: suff. stat. of μ

- The information in $\{T_1, \cdots, T_n \mid \bar{T}\}$ is equivalent to that in $\{T_1-\bar{T}, \cdots, T_n-\bar{T}\}$ — contain only information about σ^2
- If μ: nuisance parameter, conditioning on suff. stat. of μ is a good way to "throw out" useless information in data.



  - Conditional probability of the observed data in jth matched set

conditional pmf of z_{ij}'s $\mid S_j$ → likelihood

$$P(z_{0j}=1, z_{1j}=0, \ldots, z_{Mj}=0 \mid S_j = 1)$$

∵ MCCD & $S_j=1$; suff. stat. of α_j

$$= P(z_{0j}=1, z_{1j}=0, \ldots, z_{Mj}=0)/P(S_j=1)$$

$$= \frac{\cancel{\alpha_j}\ exp(\boldsymbol{x}_{0j}^T\beta)}{\sum_{i=0}^M exp(\boldsymbol{x}_{ij}^T\beta)\ \cancel{\alpha_j}} = \frac{exp(\boldsymbol{x}_{0j}^T\beta)}{exp(\boldsymbol{x}_{0j}^T\beta) + \sum_{i=1}^M exp(\boldsymbol{x}_{ij}^T\beta)}$$

$$= \left\{1 + \textstyle\sum_{i=1}^M exp\left[(\boldsymbol{x}_{ij} - \boldsymbol{x}_{0j})^T\beta\right]\right\}^{-1}$$

reference of x in jth block

data from MCCD always has this form (Note. In data collection, z_{ij}'s are fixed; in data analysis, z_{ij}'s are treated as random)

$$p_{ij} = \frac{\exp(\alpha_j + x_{ij}^T\beta)}{1+\exp(\alpha_j + x_{ij}^T\beta)}, \quad 1-p_{ij} = \frac{1}{1+\exp(\cdots)}$$

$P(Z_{0j}=1, Z_{1j}=0, \cdots, Z_{Mj}=0)$
$+ P(Z_{0j}=0, Z_{1j}=1, \cdots, Z_{Mj}=0)$
$+ \cdots$
$+ P(Z_{0j}=0, Z_{1j}=0, \cdots, Z_{Mj}=1)$

$= P(Z_{0j}=1)P(Z_{1j}=0)\times\cdots\times P(Z_{Mj}=0)$ ← independent
$= p_{0j}(1-p_{1j})\times\cdots\times(1-p_{Mj})$
$= [\exp(\alpha_j + x_{0j}^T\beta)\times 1\times\cdots\times 1] \Big/ \prod_{i=0}^{M}[1+\exp(\alpha_j + x_{ij}^T\beta)]$

➢ The complete conditional likelihood = P(all z_{ij}'s | all S_j's)

Note. nuisance parameters α_j's are gone in this likelihood

$$\mathcal{L}(\beta) = \prod_{j=1}^{n} \left\{ 1 + \sum_{i=1}^{M} exp\left[(\boldsymbol{x}_{ij} - \boldsymbol{x}_{0j})^T \beta \right] \right\}^{-1}$$

∵ data from different matched sets are assumed independent

cf. profile likelihood in LNp.3-13

cf. exact logistic in LNp.3-31

- Standard likelihood methods can now be employed to make inference
  - estimation of β ⇒ MLE
  - testing (& C.I.) of β ⇒ likelihood ratio, Wald, ...
- The conditional likelihood takes the same form as that used for the proportional hazards model (PHM) in survival/reliability analysis

lifetime (>0) data: $T_1, T_2, \cdots, T_n$ ↕ x_1 x_2 ... x_n

hazard function at x: $h_x(t) = h_0(t) \times exp(x^T \beta)$

reference hazard: nonparametric modeling (parametric modeling)

⇒ can use software developed for PHM

- Note that in this approach: $h_{x_2}(t) / h_{x_1}(t) = exp[(x_2 - x_1)^T \beta]$

within block comparison

- ∵ α_j's are not estimated ⇒ cannot make predictions of p_{ij}'s (Even if α_j's are estimated, they are not likely the estimates of the true effects of **W**, prediction of p_{ij}'s is still questionable)

use unconditional approach

check ★ in LNp.3-44: $P(Z=1 | \mathbf{W}=w_j, X=x_{ij})$

∵ α_j's in MCCD usually not the true block effects of **W** (check (Δ) in LNp.3-44)

- only can make statements about the odds ratio as measured by the β

$O_{ij}/O_{i'j} = exp[(x_{ij} - x_{i'j})^T \beta]$ as x change from x_{ij} to $x_{i'j}$

cf. $O_{ij}/O_{i'j}$ is irrelevant to **W** (α_j's) (∵ assume main-effect model, LNp.3-44)

Note. $\mathbf{W} = w_j$ is fixed

❖ **Reading**: Faraway (2006, 1st ed.), 2.12