Part IV:

Marginal models

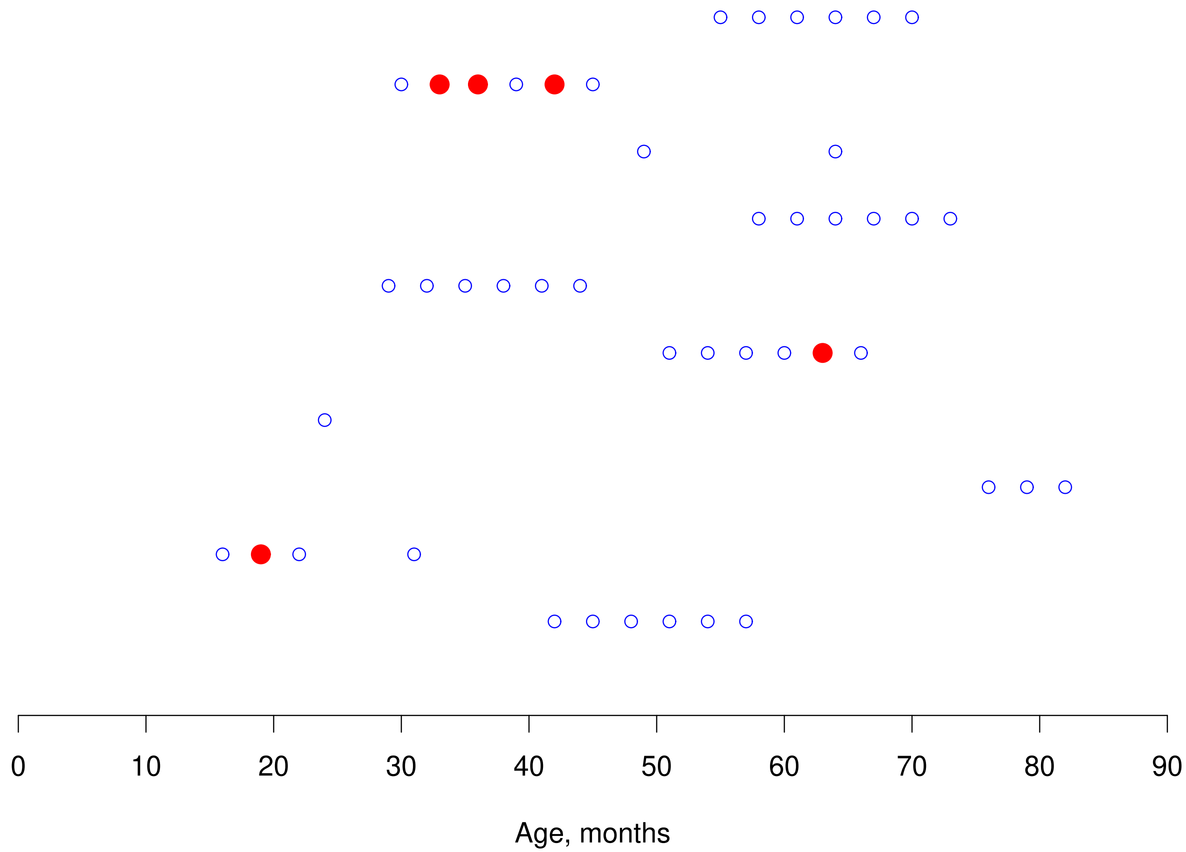
Indonesian Children's Health Study (ICHS)

- Study conducted to determine the effects of vitamin A deficiency in preschool children
- Data on $K\!=\!275$ children examined at up to six visits
  - ⋆ all pre-schoolers
  - ⋆ data and documentation available on the course website
- Primary outcome is presence/absence of a respiratory infection
  - ⋆ binary response that varies over time
- Primary exposure is xerophthalmia
  - ⋆ an ocular manifestation of vitamin A deficiency
  - ⋆ serves as a surrogate since vitamin A intake was not measured
  - ⋆ also varies over time

```
>
> load("ICHS.RData")
>
> ichs[1:10,]
      id infection age xerop cost sint gender hfora time
1  121013         0  31     0   -1    0      0    -3    1
2  121013         0  34     0    0   -1      0    -3    2
...
6  121013         0  46     0    0   -1      0    -3    6
7  121113         0  -9     0   -1    0      1     2    1
8  121113         0  -6     0    0   -1      1     0    2
...
> ##
>
> table(table(ichs$id))

  1   2   3   4   5   6
 22  32  29  55  15 122
>
> summary(ichs$age+36)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    4.0    21.0    39.0    39.2    55.0    86.0
```

- Visualization of the response for a random sample of 10 children
  - ⋆ blue circles indicate the absence of an infection
  - ⋆ red dots indicate the presence of an infection



  - ⋆ note that children have heterogeneous visit schedules

- As we consider these data, we might be interested in modeling the mean response as a function of covariates, particularly the presence/absence of xerophthalmia

$$\mu_{ki} \;=\; \mathsf{E}[Y_{ki} | \; \boldsymbol{X}_{ki}] \;=\; \Pr(Y_{ki} = 1 | \; \boldsymbol{X}_{ki})$$

- In Parts II and III we considered linear models for dependent continuous response data
  - ⋆ e.g. length in the dental growth
  - ⋆ e.g. CD4 count in the MACS data

- As we move beyond continuous response data we need a framework that
  - ⋆ permits modeling means that take on values on a restricted range
  - ⋆ simultaneously provides valid inference in the presence of
    - ∗ mean-variance relationships
    - ∗ dependence between study units

Generalized linear models

- In Methods I, you moved beyond continuous response data by considering the *generalized linear models* (GLM) framework

- GLMs are a class of parametric statistical models for the conditional distribution of a response Y_i given a p-vector of covariates X_i:

  (1) probability distribution, $Y_i \sim f_Y(y)$

  (2) linear predictor, $\eta_i = X_i^T \boldsymbol{\beta}$

  (3) link function, $g(\cdot)$ such that $g(\mu_i) = \eta_i$

- The theory of GLMs focuses on distributions that belong to the *exponential dispersion family*

$$f_Y(y_i; \theta_i, \phi) \;=\; \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} \;+\; c(y_i, \phi) \right\}$$

- Given an i.i.d sample of size n, the log-likelihood is

logit = expit^-1

$$\ell(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \frac{Y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(Y_i, \phi)$$

from which, using the chain rule, the score for β_j is:

$$\frac{\partial \ell(\boldsymbol{\beta}, \phi)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \eta_i} \frac{X_{j,i}}{V(\mu_i) a_i(\phi)} (Y_i - \mu_i)$$

- Re-writing this expression, the score for $\boldsymbol{\beta}$ can be written as:

$$\begin{aligned} \boldsymbol{U}_n(\boldsymbol{\beta}; \phi) &= \sum_{i=1}^{n} \boldsymbol{U}(\boldsymbol{\beta}; Y_i, \phi) \\ &= \sum_{i=1}^{n} D_i^T V_i^{-1} (Y_i - \mu_i) \end{aligned}$$

where $D_i = \partial \mu_i / \partial \boldsymbol{\beta}$ and $V_i = V(\mu_i) a_i(\phi)$.

we may need dispersion in count data, for binary data dispersion equals 1

Quasi-likelihood

- When the response is binary or polytomous, there is only one possible distribution and one can only proceed via maximum likelihood
  - ⋆ still, of course, substantial flexibility in how one specifies the mean of that distribution

- If the response is not binary, there are many possible distributions
  - ⋆ e.g. for continuous responses
  - ⋆ e.g. for count responses

- In some settings it may be undesirable to have to specify the full probability distribution for the data
  - ⋆ appealing properties of likelihood-based estimators (i.e. the MLE) rely on the specification of the probability distribution for the data being correct

- In quasi-likelihood one proceeds by solely specifying the first two moments
- For example, one could specify:

$$\begin{aligned} \mathsf{E}[Y_i | \boldsymbol{X}_i] &= \mu_i = g^{-1}(X_i^T \boldsymbol{\beta}) \\ \mathsf{V}[Y_i | \boldsymbol{X}_i] &= \phi V(\mu_i) \end{aligned}$$

  - ⋆ variance is the combination of some function of the mean and an additional dispersion parameter, ϕ
- Estimation/inference can then proceed on the basis of the *quasi log-likelihood*

$$\ell_q(\boldsymbol{\beta}, \phi) = \sum_{i=1}^{n} \frac{y_i \theta_i - b(\theta_i)}{\phi}$$

  - ⋆ we have (artificially) set $a_i(\phi) = \phi$ and $c(y_i, \phi) = 0$
  - ⋆ ‘quasi’ in the sense that it doesn’t correspond to any particular distribution

- Taking derivatives, the quasi-score for $\boldsymbol{\beta}$ can be written as:

$$\begin{aligned}\widetilde{\boldsymbol{U}}_n(\boldsymbol{\beta};\phi) &= \sum_{i=1}^{n} \widetilde{\boldsymbol{U}}(\boldsymbol{\beta};Y_i,\phi) \\ &= \sum_{i=1}^{n} D_i^T \widetilde{V}_i^{-1}(Y_i-\mu_i)\end{aligned}$$

where $D_i = \partial\mu_i/\partial\boldsymbol{\beta}$ and $\widetilde{V}_i = V(\mu_i)\phi$

- Note, the only difference between the quasi-score and the likelihood-based score

$$\boldsymbol{U}_n(\boldsymbol{\beta};\phi) = \sum_{i=1}^{n} D_i^T V_i^{-1}(Y_i-\mu_i)$$

is in the distinction between V_i and $\widetilde{V}_i$:

$$\begin{aligned} V_i &= V(\mu_i)a_i(\phi) && \text{likelihood-based score} \\ \widetilde{V}_i &= V(\mu_i)\phi && \text{quasi-score}\end{aligned}$$

Unbiased estimating equations

- The equations that result from setting $\boldsymbol{U}_n(\boldsymbol{\beta}; \phi)$ and $\widetilde{\boldsymbol{U}}_n(\boldsymbol{\beta}; \phi)$ to zero belong to a broader class of *unbiased estimating equations*

- Suppose Y_1, Y_2, ... are i.i.d and that $g(y; \boldsymbol{\theta})$ is some continuously differentiable function of the data such that:

$$\mathsf{E}_{\boldsymbol{\theta}}[g(Y; \boldsymbol{\theta})] \;=\; \mathbf{0}, \quad \forall\, \boldsymbol{\theta} \tag{1}$$

- Write

$$h_n(\boldsymbol{\theta}) \;=\; \frac{1}{n}\sum_{i=1}^{n} g(Y_i; \boldsymbol{\theta}) \qquad \text{e.g. sample mean} \tag{2}$$

- Refer to expression (2) as an *estimating function* or (more typically) a set of *estimating functions*

- Setting $h_n(\boldsymbol{\theta}) = \mathbf{0}$ defines a set of *estimating equations*

- Note, $\boldsymbol{U}_n(\boldsymbol{\beta};\phi)$ and $\widetilde{\boldsymbol{U}}_n(\boldsymbol{\beta};\phi)$ have the same form as (2) with

$$g(Y_i;\boldsymbol{\theta}) \;=\; \begin{cases} \boldsymbol{U}(\boldsymbol{\beta};Y_i,\phi) \;=\; D_i^T V_i^{-1}(Y_i-\mu_i) & \text{likelihood-based score} \\ \widetilde{\boldsymbol{U}}(\boldsymbol{\beta};Y_i,\phi) \;=\; D_i^T \widetilde{V}_i^{-1}(Y_i-\mu_i) & \text{quasi-score} \end{cases}$$

(if the mean model is correctly specified, quasi-score has zero mean)

- We say the estimating equations are *unbiased* if

$$\mathsf{E}_{\boldsymbol{\theta}}[h_n(\boldsymbol{\theta})] \;=\; \mathbf{0}, \quad \forall\, \boldsymbol{\theta}$$

  - ⋆ follows directly if (1) holds

- Note, the likelihood-based and quasi-likelihood-based estimating equations for GLMs are both unbiased as long as the mean is correctly specified
  - ⋆ i.e. if $\mathsf{E}[Y_i|\,\boldsymbol{X}_i] = \mu_i$ then

$$\mathsf{E}[\boldsymbol{U}_n(\boldsymbol{\beta};\phi)] \;=\; \mathsf{E}[\widetilde{\boldsymbol{U}}_n(\boldsymbol{\beta};\phi)] \;=\; \mathbf{0}$$

- Given an i.i.d sample of size n, setting $h_n(\boldsymbol{\theta})$ equal to zero and solving defines an estimator:

$$h_n(\widehat{\boldsymbol{\theta}}) \;=\; \mathbf{0}$$

- Performing a Taylor series expansion around the true value, $\boldsymbol{\theta}_0$

$$\mathbf{0} \;= h_n(\widehat{\boldsymbol{\theta}}) \;=\; h_n(\boldsymbol{\theta}_0) \;+\; \frac{\partial h_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \;+\; o_p(n^{-1/2})$$

which can be re-written as:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \;=\; \left[-\frac{\partial h_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right]^{-1} \left[\sqrt{n} h_n(\boldsymbol{\theta}_0)\right] \;+\; o_p(1)$$

- From the Central Limit Theorem:

$$\sqrt{n} h_n(\boldsymbol{\theta}_0) \;\longrightarrow\; \text{Normal}(\mathbf{0},\; \boldsymbol{\mathcal{I}})$$

where $\boldsymbol{\mathcal{I}} = \mathsf{V}_{\boldsymbol{\theta}_0}[g(Y_i; \boldsymbol{\theta}_0)]$

- Furthermore, from the Law of Large Numbers:

$$-\frac{\partial h_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \longrightarrow \boldsymbol{\mathcal{F}}$$

  where $\boldsymbol{\mathcal{F}} = -\mathsf{E}_{\boldsymbol{\theta}_0}[\partial g(Y_i;\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}]$

- By Slutsky's Theorem, we therefore have:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \longrightarrow \text{Normal}(\mathbf{0},\ \boldsymbol{\mathcal{F}}^{-1}\boldsymbol{\mathcal{I}}(\boldsymbol{\mathcal{F}}^{-1})^T)$$

- Hence, returning to the likelihood-based and quasi-likelihood-based estimators for GLMs, we can use this result to:
  (1) see that each estimator converges to the true value $\boldsymbol{\beta}_0$ as $n \to \infty$
  (2) derive the appropriate 'sandwich' form of the asymptotic variance

- Note, ϕ remains unknown and would need to be estimated
  - ⋆ simultaneously for maximum likelihood
  - ⋆ separately for quasi-likelihood

Q: If both sets of estimating equations are unbiased, what are the roles of $D_i^T V_i^{-1}$ and $D_i^T \widetilde{V}_i^{-1}$?

- ⋆ the derivative $D_i = \partial\mu_i / \partial\boldsymbol{\beta}$ serves to change the space that is being focused on when we compare ‘observed’ and ‘expected’ outcomes:
  - ∗ $\mu_i \Rightarrow \boldsymbol{\beta}$
- ⋆ V_i^{-1} and $\widetilde{V}_i^{-1}$ serve as a weights
  - ∗ from Gauss-Markov, the ‘optimal’ weighting strategy is to use the inverse of $\mathsf{V}[Y_i]$

- Hence the main distinction between the two strategies is in their efficiency properties
  - ⋆ corresponding to the different weighting schemes

Comment

- The central appeal of quasi-likelihood is that valid estimation/inference is obtained without having to specify the full probability distribution of the data
  - ⋆ only need to specify the mean model and some variance function
  - ⋆ don't have to restrict attention to the exponential dispersion family
  - ⋆ valid inference is 'guaranteed' via the robust sandwich estimator, in large samples at least

- Referred to as a *semi-parametric* statistical procedure

Marginal models for dependent data

- When our focus in Parts II and III was on continuous response data, we covered two strategies for accommodating dependence:
  - (i) specify a marginal mean model and separately a working covariance structure
    for consistency ... for efficiency
  - (ii) specify a conditional mean model that simultaneously structures the mean and the dependence

- Towards considering general response types, we could follow suit and build on GLMs by either:
  - ⋆ retaining the same framework for model specification and separately specify a working covariance structure
  - ⋆ modify the model specification to incorporate dependence via random effects

- In this Part we are going to focus on the former, extending quasi-likelihood to the dependent data setting by
  1. developing a framework for flexible and interpretable model specification
     * build on the GLM framework
  2. developing a framework for valid estimation/inference that accounts for the dependence structure
     * *generalized estimating equations*
- In Part V we will consider *generalized linear mixed models*

Marginal mean models

- Let $\mathsf{E}[Y_{ki}|\ \boldsymbol{X}_{ki}] = \mu_{ki}$ denote the marginal mean of the response for the i^{th} study unit in the k^{th} cluster

- In practice, one can encounter many response types that take on values a variety of ranges:

Response type	Range
Continuous	$(-\infty,\ \infty)$
Continuous	$(0,\ \infty)$
Continuous	$(0, 1)$
Binary	$\{0,\ 1\}$
Polytomous	$\{1,\ \ldots,\ J\}$
Count	$\{0,\ 1,\ \ldots,\ m\}$
Count	$\{0,\ 1,\ \ldots\}$

- Clearly, the range of values that Y_{ki} can take on will have implications for the range of values that μ_{ki} can take on

Q: How do we specify reasonable models for μ_{ki} while ensuring that we respect the appropriate range/scale of μ_{ki}?

- Achieved by constructing a linear predictor $\boldsymbol{X}_{ki}^T\boldsymbol{\beta}$ and relating it to μ_{ki} via a link function $g(\cdot)$:

$$g(\mu_{ki}) \ = \ \boldsymbol{X}_{ki}^T\boldsymbol{\beta}$$

  - ⋆ sometimes useful to use the notation $\eta_{ki} = \boldsymbol{X}_{ki}^T\boldsymbol{\beta}$

- Constructing the linear predictor follows the same general principles one uses for GLMs for independent data settings

- Given a set of covariates, $\boldsymbol{X}_{ki}$, we need to decide:
  - ⋆ which covariates to include in the model?
  - ⋆ how to include them in the model?

- For the most part, the decision of which covariates to include should be driven by scientific considerations
  - ⋆ is the goal estimation or prediction?
  - ⋆ is there a primary exposure of interest?
  - ⋆ which covariates are predictors of the response variable?
  - ⋆ are any of the covariates effect modifiers? confounders?
- In some settings, practical or data-oriented considerations may drive these decisions
  - ⋆ small sample sizes
  - ⋆ missing data
  - ⋆ measurement error/missclassification
- How one includes them in the model will also depend on a mixture of scientific and practical considerations

- In regard to the link function, $g(\cdot)$, there are often many choices

- If the response is binary, for example, specific options include:

$$
\begin{aligned}
\text{identity:} \qquad & g(\mu_{ki}) = \mu_{ki} \\
\text{log:} \qquad & g(\mu_{ki}) = \log(\mu_{ki}) \\
\text{logit:} \qquad & g(\mu_{ki}) = \log\left(\frac{\mu_{ki}}{1-\mu_{ki}}\right) \\
\text{probit:} \qquad & g(\mu_{ki}) = \text{probit}(\mu_{ki}) \\
\text{complementary log-log:} \qquad & g(\mu_{ki}) = \log\left\{-\log(1-\mu_{ki})\right\}
\end{aligned}
$$

- Typically, we choose a link function via consideration of two issues:

  (1) respect of the range of values that μ_{ki} can take

  (2) impact on the interpretability of $\boldsymbol{\beta}$

- These two issues can, of course, be at odds and one may have to contend with a trade-off between mathematical convenience and interpretability

Interpretation

- As with GLMs, the precise interpretation of components of $\boldsymbol{\beta}$ depends on a number of factors:
  - ⋆ other covariates in the model
  - ⋆ the functional form of the covariate under consideration
  - ⋆ the link function

- Notwithstanding these factors, the interpretation is *marginal* with respect to cluster membership
  - ⋆ comparison of the mean response among study units from two populations of such study units

Working covariance model

- Given the specification of a marginal mean model, the dependence structure is typically viewed as a nuisance
  - ⋆ a feature of the data that is secondary but nevertheless must be accounted for

- The strategy we are going to consider follows Part II in that we will specify some working dependence structure, which may or may not correspond to the actual dependence structure, and then develop methods that are 'robust' to this choice

- In contrast to Part II, certain response types may naturally exhibit a mean-variance relationship
  - ⋆ i.e. binary, polytomous or count data
  - ⋆ if we are going to be 'wise' about our choice of working covariance model, we should accommodate this

- General strategy is to adopt some working structure that may depend on $\boldsymbol{\mu}$ (and, therefore, $\boldsymbol{\beta}$), a dispersion parameter ϕ, and some additional correlation parameters, $\boldsymbol{\alpha}$:

$$\begin{aligned}\mathsf{V}[Y_{ki}|\ \boldsymbol{X}_k] &= \phi\, V(\mu_{ki}) \\ \boldsymbol{S}_k(\boldsymbol{\beta},\phi) &= \phi\ \mathsf{diag}\{V(\mu_{ki})\}\end{aligned}$$

$$\begin{aligned}\mathsf{Cor}[Y_{ki},Y_{kj}|\ \boldsymbol{X}_k] &= \rho_{k,ij}(\boldsymbol{\alpha}) \\ \boldsymbol{R}_k(\boldsymbol{\alpha}) &= \mathsf{matrix}\{\rho_{k,ij}(\boldsymbol{\alpha})\}\end{aligned}$$

$$\begin{aligned}\mathsf{Cov}[\boldsymbol{Y}_k|\ \boldsymbol{X}_k] &= \boldsymbol{V}_k(\boldsymbol{\beta},\boldsymbol{\alpha},\phi) \\ &= \boldsymbol{S}_k(\boldsymbol{\beta},\phi)^{1/2}\boldsymbol{R}_k(\boldsymbol{\alpha})\boldsymbol{S}_k(\boldsymbol{\beta},\phi)^{1/2}\end{aligned}$$

- As mentioned, the choice of working variance function $V(\mu_{ki})$ will typically be determined by the nature of the data
  - ⋆ e.g. $V(\mu_{ki}) = \mu_{ki}(1-\mu_{ki})$ for binary response data

- As in Part II, the choice of the working correlation matrix, $\boldsymbol{R}_k(\boldsymbol{\alpha})$, will be a function of:
  - ⋆ scientific considerations
    - ∗ i.e. substantive knowledge
  - ⋆ data considerations
    - ∗ i.e. exploratory data analysis
  - ⋆ practical considerations
    - ∗ i.e. sample size, K

- Same set of choices, including:

Working structure	$\boldsymbol{\rho_{k,ij}(\alpha)}$
Independence	0
Exchangeable	α
Auto-regressive (AR-1)	$\alpha^{\lvert i-j \rvert}$
Unstructured	α_{ij}

Comments

- Specification of the mean model, $\mu_{ki}(\boldsymbol{\beta})$, and the working covariance model, correlation model, $\boldsymbol{V}_k(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)$, does not identify a full probability model for the response $\boldsymbol{Y}_k$

- By only specifying the first two moments of the distribution of $\boldsymbol{Y}_k$, the analysis is *semi-parametric*
  - ⋆ additional assumptions would be needed to specify a full probability model and the corresponding (parametric) likelihood function

Q: Without a likelihood, how can we estimate $\boldsymbol{\beta}$ (and possibly $\boldsymbol{\alpha}$) and generate valid statistical inference that accounts for the dependence structure?

A: Construct an unbiased estimating equation and derive the asymptotic sampling distribution of the resulting estimator

- In the remainder of this part we will consider three frameworks for estimation/inference for semi-parametrically specified marginal models for arbitrary response types:

  ⋆ GEE 1.0

    ∗ Liang and Zeger (1986)

  ⋆ GEE 1.5

    ∗ Prentice (1988)
    ∗ Lipsitz, Laird and Harrington (1993)
    ∗ Yan and Fine (2004)

  ⋆ GEE 2.0

    ∗ Prentice and Zhao (1991)

- Each can be viewed as an extension of quasi-likelihood

- As we'll see, the differences between them lie in how they handle estimation of the parameters that index the (working) dependence structure (i.e. $\boldsymbol{\alpha}$)

GEE 1.0

- For a given value of $\boldsymbol{\alpha}$ and ϕ consider the estimating function:

$$\boldsymbol{U}_K(\boldsymbol{\beta};\boldsymbol{\alpha},\phi) \;=\; \sum_{k=1}^{K} \boldsymbol{U}(\boldsymbol{\beta};\boldsymbol{Y}_k,\boldsymbol{\alpha},\phi) \;=\; \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k)$$

  where

$$\begin{aligned}
\boldsymbol{\mu}_k &\equiv \boldsymbol{\mu}_k(\boldsymbol{\beta}) \;=\; g^{-1}(\boldsymbol{X}_k\boldsymbol{\beta}) \\
\boldsymbol{D}_k &\equiv \boldsymbol{D}_k(\boldsymbol{\beta}) \;=\; \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{\mu}_k(\boldsymbol{\beta}) \\
\boldsymbol{V}_k &\equiv \boldsymbol{V}_k(\boldsymbol{\beta},\boldsymbol{\alpha},\phi) \;=\; \boldsymbol{S}_k(\boldsymbol{\beta},\phi)^{1/2}\boldsymbol{R}_k(\boldsymbol{\alpha})\boldsymbol{S}_k(\boldsymbol{\beta},\phi)^{1/2}
\end{aligned}$$

- Setting $\boldsymbol{U}_K(\boldsymbol{\beta};\boldsymbol{\alpha},\phi) = \boldsymbol{0}$ defines a *generalized estimating equation*, the solution to which defines an estimator $\widehat{\beta}$

Properties

- Performing a Taylor series expansion of $\boldsymbol{U}_{K}(\widehat{\boldsymbol{\beta}};\boldsymbol{\alpha},\phi)$ around the true $\boldsymbol{\beta}_0$, and using the same arguments on slides 258-259, one can show that:

$$\sqrt{K}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) \ \longrightarrow \ \mathsf{MVN}_p\left(\mathbf{0}, \lim_{K\to\infty} K\boldsymbol{C}_K\right)$$

where $\boldsymbol{C}_K = \boldsymbol{A}_K^{-1}\boldsymbol{B}_K\boldsymbol{A}_K^{-1}$ with

$$\begin{aligned}
\boldsymbol{A}_K &= \sum_{k=1}^{K}\boldsymbol{D}_k^T\boldsymbol{V}_k^{-1}\boldsymbol{D}_k \\
\boldsymbol{B}_K &= \sum_{k=1}^{K}\mathsf{E}[\boldsymbol{U}(\boldsymbol{\beta};\boldsymbol{Y}_k,\boldsymbol{\alpha},\phi)\boldsymbol{U}(\boldsymbol{\beta};\boldsymbol{Y}_k,\boldsymbol{\alpha},\phi)^T] \\
&= \sum_{k=1}^{K}\boldsymbol{D}_k^T\boldsymbol{V}_k^{-1}\mathsf{Cov}[\boldsymbol{Y}_k]\boldsymbol{V}_k^{-1}\boldsymbol{D}_k
\end{aligned}$$

- $\widehat{\boldsymbol{\beta}}$ is a consistent estimator for $\boldsymbol{\beta}$ even if the working covariance model does not correspond to the true covariance model
  - ⋆ i.e. $\widehat{\boldsymbol{\beta}}$ is 'robust' to misspecification of the dependence structure (in large samples, at least)

- Selecting a working covariance model that is 'close' to the true covariance model will yield a more efficient estimator
  - ⋆ alternative choices correspond to alternative weighting schemes
  - ⋆ the optimal weighting scheme is to use the true covariance model

- Regardless of the true covariance model, valid inference can be performed with the use of $\boldsymbol{C}_K = \boldsymbol{A}_K^{-1} \boldsymbol{B}_K \boldsymbol{A}_K^{-1}$
  - ⋆ since $\boldsymbol{B}_K$ depends on the unknown true covariance model, we typically plug in an empirical estimate of it and base inference on:

$$\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}] \;=\; \boldsymbol{A}_K^{-1} \left(\sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k)(\boldsymbol{Y}_k - \boldsymbol{\mu}_k)^T \boldsymbol{V}_k^{-1} \boldsymbol{D}_k \right) \boldsymbol{A}_K^{-1}$$

Estimation of $(\boldsymbol{\alpha}, \phi)$

- While the asymptotic and robustness properties of $\widehat{\boldsymbol{\beta}}$ hold regardless of the specific working structure adopted, operationally we need to plug something in for the unknown $(\boldsymbol{\alpha}, \phi)$

- Liang and Zeger (1986) propose to proceed by plugging in simple moment-based estimators based on the scaled/standardized residuals

- For example, if ϕ requires estimation, one could use:

$$\widehat{\phi} \;=\; \frac{1}{N-p} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \frac{(Y_{ki} - \hat{\mu}_{ki})^2}{V(\hat{\mu}_{ki})}$$

where $\hat{\mu}_{ki} \;\equiv\; \mu_{ki}(\widehat{\boldsymbol{\beta}}) \;=\; g^{-1}(\boldsymbol{X}_k \widehat{\boldsymbol{\beta}})$

- Moment-based estimators for $\boldsymbol{\alpha}$ depend on the working covariance model that is adopted

- For example, if a working exchangeable structure is adopted, one can use:

$$\widehat{\alpha} = \frac{1}{N^* - p} \sum_{k=1}^{K} \sum_{i<j} R_{ki} R_{kj}$$

where $N^* = \sum_k n_k(n_k - 1)/2$ and

$$R_{ki} = \frac{Y_{ki} - \hat{\mu}_{ki}}{\sqrt{\phi V(\hat{\mu}_{ki})}}$$

$\star$ plug in an estimate of ϕ as needed

Theorem 2 (Liang and Zeger, 1986)

- Under mild regularity conditions and if
  - ⋆ $\widehat{\phi}$ is $\sqrt{K}$-consistent, given $\boldsymbol{\beta}$
  - ⋆ $\widehat{\boldsymbol{\alpha}}$ is $\sqrt{K}$-consistent, given $\boldsymbol{\beta}$ and ϕ
  - ⋆ $|\partial\boldsymbol{\alpha}(\boldsymbol{\beta},\phi)/\partial\phi|$ is bounded

  then

$$\widehat{\boldsymbol{\beta}} \;\sim\; \mathrm{MVN}_p(\boldsymbol{\beta}_0,\, \boldsymbol{C}_K)$$

- Note, these assumptions may not be satisfied with certain moment-based estimators for the covariance model parameters if the correlation structure is incorrectly specified
  - ⋆ Crowder (Biometrika, 1995)
  - ⋆ motivates using alternative strategies for estimating $\boldsymbol{\alpha}$

Computation

- To summarize, GEE 1.0 involves solving an estimating equation to obtain an estimate of $\boldsymbol{\beta}$ for a given value of $(\boldsymbol{\alpha}, \phi)$:

$$\mathbf{0} \;=\; \sum_{k=1}^{K} \boldsymbol{D}_k(\boldsymbol{\beta})^T \boldsymbol{V}_k(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k(\boldsymbol{\beta}))$$

  while using moment-based estimators for $(\boldsymbol{\alpha}, \phi)$

- Operationally, given an initial estimate of $\boldsymbol{\beta}$ one can proceed by iterating between two steps until convergence is acheived:
  1. Given $\boldsymbol{\beta}$, calculate $(\widehat{\boldsymbol{\alpha}}, \widehat{\phi})$
  2. Given $(\widehat{\boldsymbol{\alpha}}, \widehat{\phi})$, update the current estimate of $\boldsymbol{\beta}$ via Fisher scoring

$$\widehat{\boldsymbol{\beta}}^{(s+1)} \;=\; \widehat{\boldsymbol{\beta}}^{(s)} \;+\; \left\{ \sum_{k=1}^{K} \left(\boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{D}_k \right)^{-1} \left(\boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \right) \right\}$$

  with the ‘update’ term evaluated at $(\widehat{\boldsymbol{\beta}}^{(s)}, \widehat{\boldsymbol{\alpha}}, \widehat{\phi})$.

Inference

- Consider testing the linear hypotheses of the form:

$$\mathrm{H}_0 : \boldsymbol{Q}\boldsymbol{\beta} \;=\; \mathbf{0}$$

where $\boldsymbol{Q}$ is a matrix of full rank with $\dim(\boldsymbol{Q}) = r \times p$ with $r < p$

- Given the GEE estimate $\widehat{\boldsymbol{\beta}}$, one can evaluate evidence regarding the null on the basis of the multivariate Wald statistic:

$$(\boldsymbol{Q}\widehat{\boldsymbol{\beta}})^T(\boldsymbol{Q}\widehat{\mathsf{Cov}}[\widehat{\boldsymbol{\beta}}]\boldsymbol{Q}^T)^{-1}(\boldsymbol{Q}\widehat{\boldsymbol{\beta}}) \;\sim\; \chi^2_r$$

where $\widehat{\mathsf{Cov}}[\widehat{\boldsymbol{\beta}}]$ is the empirical version of the sandwich estimator

$$\widehat{\mathsf{Cov}}[\widehat{\boldsymbol{\beta}}] \;=\; \boldsymbol{A}_K^{-1}\left(\sum_{k=1}^{K}\boldsymbol{D}_k^T\boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)(\boldsymbol{Y}_k-\boldsymbol{\mu}_k)^T\boldsymbol{V}_k^{-1}\boldsymbol{D}_k\right)\boldsymbol{A}_K^{-1}$$

- Consider testing the hypotheses:

$$\mathrm{H}_0 : \begin{bmatrix} \boldsymbol{\beta}_1 \\ \mathbf{0} \end{bmatrix} \quad \text{versus} \quad \mathrm{H}_0 : \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}$$

with $\dim(\boldsymbol{\beta}_2) = r$

- One can evaluate evidence regarding the null on the basis of the score statistic:

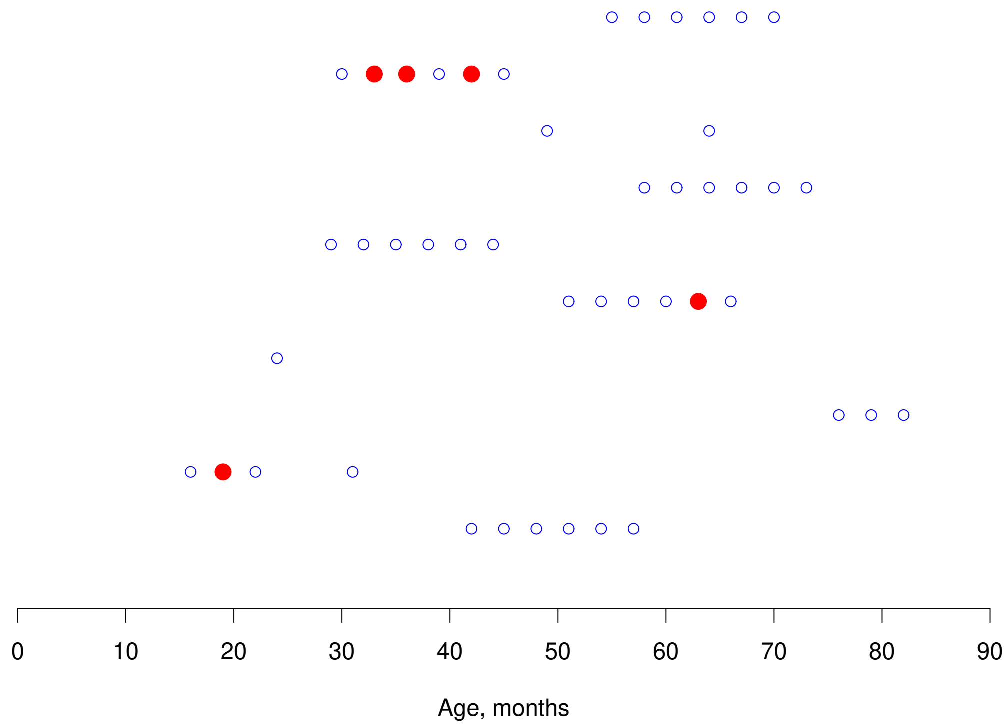
$$\frac{1}{K}(\boldsymbol{U}_K(\widehat{\boldsymbol{\beta}}_0; \widehat{\boldsymbol{\alpha}}_0, \hat{\phi}_0))^T (\widehat{\mathsf{Cov}}[\widehat{\boldsymbol{\beta}}_0])^{-1} (\boldsymbol{U}_K(\widehat{\boldsymbol{\beta}}_0; \widehat{\boldsymbol{\alpha}}_0, \hat{\phi}_0)) \ \sim \ \chi^2_r$$

where $\widehat{\boldsymbol{\beta}}_0 = (\widehat{\boldsymbol{\beta}}_1, \mathbf{0})$ denotes the GEE estimator obtained under the null, $(\widehat{\boldsymbol{\alpha}}_0, \widehat{\phi}_0)$ are the corresponding moment-based estimates and $\widehat{\mathsf{Cov}}[\widehat{\boldsymbol{\beta}}_0]$ is the empirical covariance evaluated at $\widehat{\boldsymbol{\beta}}_0$

- Note, since no explicit likelihood is used the likelihood ratio test is not available

Application to the ICHS data

- Primary interest lies with the binary response Y_{ki}=0/1=absence/presence of a respiratory infection
  - ⋆ longitudinal response data for a sample of 10 children from the K=275 in the study



- Recall that presence of xerophthalmia, an ocular manifestation of vitamin A deficiency, is the primary exposure of interest

- Consider the model:

$$\text{logit}\,\Pr(Y_{ki}|\ \boldsymbol{X}_{ki}) \;=\; \beta_0 \;+\; \beta_1\text{xerophthalmia}_{ki} \;+\; \beta_2\text{Age}_{ki}$$

  ⋆ the odds ratio $\exp\{\beta_1\}$ can be interpreted as the relative difference in the odds of a respiratory infection between two populations of children of the same age but that differ in that one is vitamin A deficient and the other not

- Note, $\exp\{\beta_1\}$ is a *population-averaged* or *marginal* parameter

- To estimate and perform inference for $\boldsymbol{\beta}$ in this model one can use the `gee()` function in the gee package

```
>
> library(gee)
> ?gee
```

- The basic call to `gee()` has the following key elements:

`formula`	Specification for the marginal mean model, $\boldsymbol{X}_k\boldsymbol{\beta}$
`id`	Variable that indicates cluster membership
`scale.fix`	Indicator of whether ϕ should be fixed (default = F)
`family`	Family object as in `glm()`
`corstr`	Working correlation structure (default = "independence")
`Mv`	Required for "stat_M_dep", "non_stat_M_dep", and "AR-M" structures (default = 1)

- Note, `corstr` is a character string with the following additional options:
  ⋆ "fixed", "stat_M_dep", "non_stat_M_dep", "exchangeable", "AR-M", "unstructured"

- The "stat_M_dep" and "non_stat_M_dep" options correspond to stationary and non-stationary banded matrixes

```
> ## Working independence
> ##
> fit0.gee <- gee(infection ~ xerop + age, id=id, data=ichs,
+                 family=binomial, scale.fix=TRUE, corstr="independence")
Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
running glm to get initial regression estimate
(Intercept)       xerop         age
-2.38479528  0.72015484 -0.02605769
>
> ## Working exchangeable
> ##
> fit1.gee <- gee(infection ~ xerop + age, id=id, data=ichs,
+                 family=binomial, scale.fix=TRUE, corstr="exchangeable")
...
>
> ## Working auto-regressive with Mv=1
> ##
> fit2.gee <- gee(infection ~ xerop + age, id=id, data=ichs,
+                 family=binomial, scale.fix=TRUE, corstr="AR-M")
...
Error in gee(infection ~ xerop + age, id = id, data = ichs, family = binomial,  :
  cgee: M-dependence, M=1, but clustsize=1
fatal error for this model
```

- It's unclear what is going on with `gee()` and the auto-regressive structure, so we can use `geeglm()` from the `geepack` packages

```
> ##
> library(geepack)
> ?geeglm
>
> ## Working independence
> ##
> fit0.pack <- geeglm(infection ~ xerop + age, id=id, data=ichs,
+                     family=binomial, scale.fix=TRUE, corstr="independence")
>
> ## Working exchangeable
> ##
> fit1.pack <- geeglm(infection ~ xerop + age, id=id, data=ichs,
+                     family=binomial, scale.fix=TRUE, corstr="exchangeable")
>
> ## Working auto-regressive with Mv=1
> ##
> fit2.pack <- geeglm(infection ~ xerop + age, id=id, data=ichs,
+                     family=binomial, scale.fix=TRUE, corstr="ar1")
```

  - ⋆ success!

```
> summary(fit0.gee)
...
 Link:                      Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Independent
...
Coefficients:
               Estimate  Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -2.38479528 0.110049742 -21.670158 0.117276627 -20.334787
xerop        0.72015485 0.428114719   1.682154 0.419677792   1.715971
age         -0.02605769 0.005675641  -4.591145 0.005289622  -4.926191

Estimated Scale Parameter:  1
Number of Iterations:  1

Working Correlation
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1    0    0    0    0    0
[2,]    0    1    0    0    0    0
[3,]    0    0    1    0    0    0
[4,]    0    0    0    1    0    0
[5,]    0    0    0    0    1    0
[6,]    0    0    0    0    0    1
```

```
> summary(fit1.gee)
...
 Link:                       Logit
 Variance to Mean Relation: Binomial
 Correlation Structure:     Exchangeable
...
Coefficients:
               Estimate  Naive S.E.    Naive z Robust S.E.   Robust z
(Intercept) -2.37017134 0.118038560 -20.079636  0.11680055 -20.292467
xerop        0.58774620 0.449417260   1.307796  0.44974026   1.306857
age         -0.02532606 0.006047399  -4.187926  0.00525355  -4.820751

Estimated Scale Parameter:  1
Number of Iterations:  3

Working Correlation
           [,1]       [,2]       [,3]       [,4]       [,5]       [,6]
[1,] 1.00000000 0.04418081 0.04418081 0.04418081 0.04418081 0.04418081
[2,] 0.04418081 1.00000000 0.04418081 0.04418081 0.04418081 0.04418081
[3,] 0.04418081 0.04418081 1.00000000 0.04418081 0.04418081 0.04418081
[4,] 0.04418081 0.04418081 0.04418081 1.00000000 0.04418081 0.04418081
[5,] 0.04418081 0.04418081 0.04418081 0.04418081 1.00000000 0.04418081
[6,] 0.04418081 0.04418081 0.04418081 0.04418081 0.04418081 1.00000000
```

- Output from `geeglm()` is slightly different
  - ⋆ only returns the robust standard error estimate
  - ⋆ returns measures of uncertainty for α when they are estimated
  - ⋆ reflects a different approach to estimation (specifically GEE 1.5)

```
> summary(fit1.pack)
...
 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept) -2.37028  0.11680 411.81  < 2e-16 ***
xerop        0.58877  0.44948   1.72     0.19
age         -0.02533  0.00525  23.25  1.4e-06 ***
...
Scale is fixed.

Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.0438  0.0277
Number of clusters:   275   Maximum cluster size: 6
```

```
> summary(fit2.pack)
...
 Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept) -2.37478  0.11694 412.43  < 2e-16 ***
xerop        0.63749  0.44352   2.07     0.15
age         -0.02598  0.00527  24.32  8.1e-07 ***
...
Scale is fixed.

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.0581  0.0401
Number of clusters:   275   Maximum cluster size: 6
```

- Summary of results:

	gee()			geeglm()	
	Est	SE_{model}	SE_{robust}	Est	SE_{robust}
Working independence					
xerophthalmia	0.720	0.428	0.420	0.720	0.420
Age, years	-0.026	0.007	0.005	-0.026	0.005
Working exchangeable					
xerophthalmia	0.588	0.449	0.450	0.589	0.449
Age, years	-0.025	0.006	0.005	-0.025	0.005
ρ	0.044			0.044	0.028
Working AR-1					
xerophthalmia				0.638	0.443
Age, years				-0.026	0.005
ρ				0.058	0.040

- No evidence of a relationship between xerophthalmia and risk of respiratory infection
  - ⋆ at least given the (very) simple mean model that is adopted
- Extent of dependence among observations for any given child appears to be small
- The similarity between the model-based and robust standard errors for the estimates based on a working exchangeable correlation structure suggests that this might be a reasonable approximation to the true dependence structure

Comments

- Moment-based estimation of $\boldsymbol{\alpha}$ has its drawbacks
  - ⋆ can be inefficient, relative to model-based estimation
  - ⋆ only offers a limited class of dependence structures
  - ⋆ attention restricted (primarily) to correlation as a measure of dependence, which may not be well-suited for all data types

Q: Can we improve upon moment-based estimation for $\boldsymbol{\alpha}$?

A: Yes!

- ⋆ consider estimators of $\boldsymbol{\alpha}$ that are more efficient but do not sacrifice robustness in estimation of $\boldsymbol{\beta}$
  - ∗ GEE 1.5
- ⋆ consider estimators that jointly target $(\boldsymbol{\beta}, \boldsymbol{\alpha})$, and are efficient for both
  - ∗ GEE 2.0

GEE 1.5

- Focusing on binary response data, the central idea of GEE 1.5 is to replace the moment-based estimators of $\boldsymbol{\alpha}$ with a different estimator based on an additional set of estimating equations
  - ⋆ Prentice (Biometrika, 1988)

- For a given marginal mean model, $\mathsf{E}[Y_{ki}] = \mu_{ki} = \mu_{ki}(\boldsymbol{\beta})$, let

$$Z_{k,ij} \equiv Z_{k,ij}(\boldsymbol{\beta}) = \frac{(Y_{ki} - \mu_{ki})(Y_{kj} - \mu_{kj})}{\sqrt{\mu_{ki}(1 - \mu_{ki})}\sqrt{\mu_{kj}(1 - \mu_{kj})}}$$

- If we let $\rho_{k,ij} = \mathsf{Cor}(Y_{ki}, Y_{kj})$, then one can show that:

$$\begin{aligned}
\mathsf{E}[Z_{k,ij}] &= \rho_{k,ij} \\
\mathsf{V}[Z_{k,ij}] &= 1 + \rho_{k,ij} \frac{(1 - 2\mu_{ki})(1 - 2\mu_{kj})}{\sqrt{\mu_{ki}(1 - \mu_{ki})}\sqrt{\mu_{kj}(1 - \mu_{kj})}} - \rho_{k,ij}^2
\end{aligned}$$

- Hence, given a mean model specified as a function of $\boldsymbol{\beta}$ and a working covariance model specified as a function of $\boldsymbol{\alpha}$, one can write down the first two moments of $Z_{k,ij}$ as functions of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$

- We can use these to form a new set of estimating equations for $\boldsymbol{\alpha}$ given $\boldsymbol{\beta}$:

$$\boldsymbol{U}_{K}(\boldsymbol{\alpha}; \boldsymbol{\beta}) \; = \; \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1} (\boldsymbol{Z}_k - \boldsymbol{\rho}_k) \; = \; \mathbf{0}$$

  where

  - ⋆ $\boldsymbol{Z}_k = (Z_{k,12}, \; Z_{k,13}, \; \ldots, \; Z_{k,23}, \; \ldots)$
    - ∗ a vector of length $n_k^* = n_k(n_k - 1)/2$
  - ⋆ $\boldsymbol{W}_k$ is an $n_k^* \times n_k^*$ working covariance matrix for $\boldsymbol{Z}_k$
  - ⋆ $\boldsymbol{E}_k = \partial \boldsymbol{\rho}_k / \partial \boldsymbol{\alpha}$ is a $n_k^* \times q$ matrix where q is the length of $\boldsymbol{\alpha}$

- Note, Prentice (1988) uses the notation $\boldsymbol{\delta}_k$ instead of $\boldsymbol{\rho}_k$

- Towards developing some intuition regarding $\boldsymbol{U}_{K}(\boldsymbol{\alpha};\boldsymbol{\beta})$, recall the estimating equations for $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$:

$$\boldsymbol{U}_{K}(\boldsymbol{\beta};\boldsymbol{\alpha}) \;=\; \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \;=\; \mathbf{0}$$

  - ⋆ $(\boldsymbol{Y}_k - \boldsymbol{\mu}_k)$ serves as the primary criterion for estimation
    - ∗ i.e. compare observed to expected under the model
  - ⋆ $\boldsymbol{V}_k$ is a working covariance matrix for $\boldsymbol{Y}_k$ that serves to inform a weighting scheme
    - ∗ impact on efficiency properties
  - ⋆ $\boldsymbol{D}_k$ serves to change the space over which the solutions are sought
    - ∗ we want estimates of $\boldsymbol{\beta}$, not $\boldsymbol{\mu}_k$

- Turning back to $\boldsymbol{U}_{K}(\boldsymbol{\alpha};\boldsymbol{\beta})$:

$$\boldsymbol{U}_{K}(\boldsymbol{\alpha};\boldsymbol{\beta}) \;=\; \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1}(\boldsymbol{Z}_k - \boldsymbol{\rho}_k) \;=\; \mathbf{0}$$

  ⋆ $(\boldsymbol{Z}_k - \boldsymbol{\rho}_k)$ serves as the primary criterion for estimation

    ∗ i.e. compare observed to expected under the model

  ⋆ $\boldsymbol{W}_k$ is a working covariance matrix for $\boldsymbol{Z}_k$ that serves to inform a weighting scheme

    ∗ impact on efficiency properties

  ⋆ $\boldsymbol{E}_k$ serves to change the space over which the solutions are sought

    ∗ we want estimates of $\boldsymbol{\alpha}$, not $\boldsymbol{\rho}_k$

Simple example

- Suppose a working exchangeable correlation structure for $\boldsymbol{Y}_k$ is adopted:

$$\boldsymbol{R}_k(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha & \ldots & \alpha \\ \alpha & 1 & \ldots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \ldots & 1 \end{bmatrix}$$

- Towards estimating α, suppose we adopt $\boldsymbol{W}_k = \boldsymbol{I}_{n_k^*}$ as a working covariance matrix for $\boldsymbol{Z}_k$
  - ⋆ takes $\mathsf{V}[Z_{k,ij}]$ to be 1 and assumes independence across the elements

- Since $\partial\rho_{k,ij}/\partial\alpha = 1 \ \forall\ k,\ i,\ j$, we can write

$$\mathbf{0} = \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1}(\boldsymbol{Z}_k - \boldsymbol{\rho}_k) = \sum_{k=1}^{K} \sum_{i<j} (Z_{k,ij} - \alpha)$$

- This expression can be solved analytically to give:

$$\hat{\alpha} \;=\; \frac{1}{N^*}\sum_{k=1}^{K}\sum_{i<j} Z_{k,ij}$$

where $N^* = \sum_k n_k(n_k-1)/2$

- Note, this estimate is very close in functional form to the moment-based estimator we used in GEE 1.0:

$$\hat{\alpha} \;=\; \frac{1}{N^*-p}\sum_{k=1}^{K}\sum_{i<j} R_{ki}R_{kj}$$

where

$$R_{ki} \;=\; \frac{Y_{ki}-\hat{\mu}_{ki}}{\sqrt{\phi V(\hat{\mu}_{ki})}}$$

$\star$ see slide 278

Choice of $\boldsymbol{W}_k$

- As with $\boldsymbol{V}_k$ in $\boldsymbol{U}_{\scriptscriptstyle K}(\boldsymbol{\beta};\boldsymbol{\alpha})$, one has considerable choice in the working covariance matrix for the $\boldsymbol{Z}_k$
  - ⋆ in contrast to $\boldsymbol{Y}_k$, however, it's not readily obvious how one goes about thinking about correlation between the elements of $\boldsymbol{Z}_k$

- One simple way forward is to use $\boldsymbol{W}_k = \boldsymbol{I}_{n_k^*}$
  - ⋆ as in the example just presented

- Another way forward is to use

$$\boldsymbol{W}_k \;=\; \text{diag}\{W_{k,12},\; W_{k,13},\; \ldots,\; W_{k,23},\; \ldots\}$$

where $W_{k,ij} = \mathsf{V}[Z_{k,ij}]$

  - ⋆ explicit acknowledgement of heteroskedasticity
  - ⋆ see slide 295 for the form of $\mathsf{V}[Z_{k,ij}]$

Paired estimating equations

- For a given specification of the first two moments for $\boldsymbol{Y}_k$ and $\boldsymbol{Z}_k$, one can estimate $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ by iterating between two sets of estimating equations:

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \\ \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1} (\boldsymbol{Z}_k - \boldsymbol{\rho}_k) \end{aligned}$$

  - ⋆ denote the solution to these equations as $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$

- Performing a Taylor series expansion of these equations around the true $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)$, and using the same arguments on slides 256-257, one can show that $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$ has an asymptotic MVN sampling distribution
  - ⋆ asymptotics are with respect to K

- The asymptotic variance-covariance matrix is again of 'sandwich' form
  - ⋆ robust to misspecification of $\boldsymbol{V}_k$ and $\boldsymbol{W}_k$
  - ⋆ see Prentice (1988) and Yan and Fine (2004)
- Note, if the marginal mean model, $\boldsymbol{\mu}_k$, is correctly specified, $\widehat{\boldsymbol{\beta}}$ is consistent regardless of whether the model for $\boldsymbol{\rho}_k$ is correctly specified
  - ⋆ interesting to think about what $\widehat{\boldsymbol{\alpha}}$ is estimating
- As in GEE 1.0, if one is confident in your specification of the model for $\boldsymbol{\rho}_k$, one could use the model-based estimate of Cov$[\widehat{\boldsymbol{\beta}}]$, instead of the sandwich-based estimates
  - ⋆ typically a simple working model for Cov$[\boldsymbol{Z}_k]$ is adopted, so that only the sandwich-based estimator for Cov$[\widehat{\boldsymbol{\alpha}}]$ would be used

Modeling $\rho_{k,ij}$

- One advantage of using estimating equations for the dependence structure is that one can easily specify a model for the pairwise correlations
  - ⋆ simplifies extending the class of choices for the working covariance matrix to depend on covariates

- Consider paired models for the means and correlations:

$$\begin{aligned} g_\mu(\mu_{ki}) &= \boldsymbol{X}_{ki}\boldsymbol{\beta} \\ g_\rho(\rho_{k,ij}) &= \widetilde{\boldsymbol{X}}_{k,ij}\boldsymbol{\alpha} \end{aligned}$$

rho is a number in [-1,1], that's why we need a link function

  - ⋆ $\widetilde{\boldsymbol{X}}_{k,ij}$ is a vector of covariates that are functions of the covariates for the i^{th} and j^{th} study units in the k^{th} cluster
  - ⋆ $g_\rho(\cdot)$ is a link function

- For example, suppose $X_{ki} = 0/1 =$ M/F is a study-unit specific covariate and consider the model:

$$\rho_{k,ij} \ = \ \alpha_0 \ + \ \alpha_1[\widetilde{X}_{k,ij} == 1] \ + \ \alpha_2[\widetilde{X}_{k,ij} == 2]$$

with

$$\widetilde{X}_{k,ij} \ = \ \begin{cases} 0 & \text{if } X_{ki} \neq X_{kj} \\ 1 & \text{if } X_{ki} = X_{kj} \ = 0 \\ 2 & \text{if } X_{ki} = X_{kj} \ = 1 \end{cases}$$

  - ⋆ α_0 is the correlation between two study units of different genders
  - ⋆ $\alpha_0 + \alpha_1$ is the correlation between two study units that are both male
  - ⋆ $\alpha_0 + \alpha_2$ is the correlation between two study units that are both female

- Beyond the identity link function, once choice that results in values on the real line is the Fisher transform:

$$g_\rho(\rho_{k,ij}) \ = \ \frac{1}{2} \log \left(\frac{1 + \rho_{k,ij}}{1 - \rho_{k,ij}} \right)$$

- However one specifies models for μ_{ki} and $\rho_{k,ij}$, one can estimate $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ via an analogous set of estimating equations:

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \\ \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1} (\boldsymbol{Z}_k - \boldsymbol{\rho}_k) \end{aligned}$$

  - ⋆ key difference is in the model specification for the components of $\boldsymbol{\rho}_k$
  - ⋆ decisions regarding $\boldsymbol{W}_k$ don't really change
  - ⋆ functional form of $\boldsymbol{E}_k$ will be different

GEE 1.5 in R

- `gee()` uses moment-based estimates for the correlation parameters

- The `geepack` package in R implements estimating equations for $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \phi)$ for a (somewhat) limited set of models
  - ⋆ `geese()` is the main fitter function that solves the estimating equations
  - ⋆ `geeglm()` is a wrapper function with a syntax that is similar to `glm()`

- A basic call to `geeglm()` is very similar to that for `gee()`:

`formula`	Specification for the marginal mean model, $\boldsymbol{X}_k\boldsymbol{\beta}$
`id`	Variable that indicates cluster membership
`scale.fix`	Indicator of whether ϕ should be fixed (default = F)
`family`	Family object as in `glm()`
`corstr`	Working correlation structure

- For certain, more complex models, a call to `geese()` will be necessary and could involve the following:

`sformula`	Specification of a model for the dispersion (scale) parameter
`scale.link`	Link function for the scale(s)
	⋆ "identity" and "log"
`zcor`	A design matrix for the correlation parameters
`cor.link`	Link function for the correlation coefficients
	⋆ "identity" and "fisherz"

ICHS data (con't)

- Consider an augmented model:

$$\begin{aligned}\text{logit}\,\Pr(Y_{ki}|\ \boldsymbol{X}_{ki}) &= \beta_0 + \beta_1\text{xerophthalmia}_{ki} \\ &\quad + \beta_2\text{Age}_{ki} \\ &\quad + \beta_3\text{Gender}_{ki}\end{aligned}$$

- Consider two working models for the dependence structure:

  (1) exchangeable with a common α for males and females

  (2) exchangeable with different α's for males and females

```
>
> ## Working exchangeable, with a common \alpha
> ##
> fit3.pack <- geeglm(infection ~ xerop + age + gender, id=id, data=ichs,
+                     family=binomial, scale.fix=TRUE,
+                     corstr="exchangeable")
```

```
> ## Define a new working correlation structure
> ##  - exchangeable but with separate \alpha for males and females
>
> ## Set-up an "empty" unstructured correlation matrix
> ##  - number of rows is the number of off-diagnonals = 2,429
> ##
> zcor <- genZcor(clusz=table(ichs$id), waves=ichs$time, corstrv=4)
>
> ## Identify which rows correspond to the males
> ##
> temp   <- table(ichs$id, ichs$gender)
> temp   <- cbind(temp, as.numeric(temp[,2] > 0))
> nkStar0 <- temp[,1] * (temp[,1] - 1) / 2
> nkStar1 <- temp[,2] * (temp[,2] - 1) / 2
> nkStar  <- pmax(nkStar0, nkStar1)
> males.expand <- rep(temp[,3], nkStar)
> rows.M <- c(1:sum(nkStar))[males.expand == 1]
>
> ## Final user-defined structure
> ##
> zcor.MF <- matrix(0, nrow(zcor), 2)
> zcor.MF[,1] <- 1
> zcor.MF[rows.M,2]  <- 1
```

```
> ## Working exchangeable with gender-specific correlation parameters
> ##
> fit4.pack <- geese(infection ~ xerop + age + gender, id=id, data=ichs,
+                    family=binomial, scale.fix=TRUE,
+                    corstr="userdefined", zcor=zcor.MF)
>
> ## RESULTS
>
> summary(fit3.pack)
...
Coefficients:
            Estimate  Std.err   Wald Pr(>|W|)
(Intercept) -2.22312  0.14880 223.21  < 2e-16 ***
xerop        0.53580  0.44525   1.45     0.23
age         -0.02510  0.00518  23.52  1.2e-06 ***
gender      -0.38658  0.23663   2.67     0.10
...
Correlation: Structure = exchangeable  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha   0.0409  0.0255
```

```
> summary(fit4.pack)
...
 Coefficients:
            estimate  san.se   wald        p
(Intercept)  -2.2184 0.14916 221.19 0.00e+00
xerop         0.4917 0.45828   1.15 2.83e-01
age          -0.0252 0.00518  23.71 1.12e-06
gender       -0.3884 0.23660   2.70 1.01e-01
...
Correlation Model:
 Correlation Structure:     userdefined
 Correlation Link:          identity

  Estimated Correlation Parameters:
        estimate san.se  wald     p
alpha:1   0.0568 0.0389 2.134 0.144
alpha:2  -0.0375 0.0485 0.596 0.440
```

- No evidence that the correlation parameter differs between the males and females

Dental growth data

- For binary data there is no dispersion parameter
- To illustrate the functionality of geese() for these parameters, let's consider (again!) the model:

$$\mathsf{E}[Y_{ki}] \ = \ \beta_0 \ + \ \beta_1 A^*_{ki} \ + \ \beta_1 G_k \ + \ \beta_3 A^*_{ki} G_k$$

- Consider various (working) specifications for:

$$\boldsymbol{V}_k \ = \ \boldsymbol{S}_k(\boldsymbol{\beta}, \phi)^{1/2} \boldsymbol{R}_k(\boldsymbol{\alpha}) \boldsymbol{S}_k(\boldsymbol{\beta}, \phi)^{1/2}$$

```
> ## Three "standard" choices
> ##
> fit0.pack <- geeglm(length ~ ageStar * gender, id=id, data=growth,
+                     corstr="independence")
> fit1.pack <- geeglm(length ~ ageStar * gender, id=id, data=growth,
+                     corstr="exchangeable")
> fit2.pack <- geeglm(length ~ ageStar * gender, id=id, data=growth,
+                     corstr="unstructured")
```

```
> ## Define a new working correlation structure
> ##  - exchangeable
> ##  - seperate correlation parameters for males and females
>
> ## Set-up an "empty" unstructured correlation matrix
> ##  - number of rows is the number of off-diagnonals = 156 = 26 x (4x(4-1)/2)
> growth$wave <- rep(1:4, 26)
> zcor <- genZcor(clusz=table(growth$id), waves=growth$wave, corstrv=4)
>
> ## Identify which rows correspond to the female
> rows.F  <- c(1:(11*2*3))
>
> ## Final user-defined structure
> zcor.MF <- matrix(0, nrow(zcor), 2)
> zcor.MF[,1]  <- 1
> zcor.MF[-rows.F,2] <- 1
>
> ## "Complex" choice (1)
> ##  - exchangeable correlation, with gender-specific parameters
> ##  - common variance on the diagonal (i.e. not gender-specific)
> ##
> fit3.pack <- geese(length ~ ageStar * gender, id=id, data=growth,
+                    corstr="userdefined", zcor=zcor.MF)
```

```
> ## "Complex" choice (2)
> ##  - exchangeable correlation, with a common parameter (i.e. not gender-specific)
> ##  - gender-specific variance on the diagonal
> ##
> fit4.pack <- geese(length ~ ageStar * gender, id=id, data=growth,
+                    sformula= ~gender, corstr="exchangeable")
>
> ## "Complex" choice (3)
> ##  - exchangeable correlation, with gender-specific parameters
> ##  - gender-specific variance on the diagonal
> ##
> fit5.pack <- geese(length ~ ageStar * gender, id=id, data=growth,
+                    sformula= ~gender, corstr="userdefined", zcor=zcor.MF)
>
> summary(fit5.pack)
...
 Coefficients:
                   estimate san.se    wald        p
(Intercept)           21.21 0.5604 1432.19 0.00e+00
ageStar                0.48 0.0631   57.70 3.05e-14
gendermale             1.49 0.7940    3.53 6.04e-02
ageStar:gendermale     0.32 0.1214    6.97 8.28e-03
```

```
Scale Model:
 Scale Link:                identity

 Estimated Scale Parameters:
            estimate san.se wald       p
(Intercept)    4.470   1.70 6.93 0.00847
gendermale     0.761   2.11 0.13 0.71800

Correlation Model:
 Correlation Structure:     userdefined
 Correlation Link:          identity

 Estimated Correlation Parameters:
        estimate san.se   wald      p
alpha:1    0.868 0.0568 233.68 0.0000
alpha:2   -0.419 0.2024   4.29 0.0384
...
```

- Suggests that the correlation parameter differs across the genders but not the variances

Dependence models via the odds ratio

- For binary responses, correlation may not be a good measure of dependence since the range of values it can take on is constrained by the marginal means

- To see this, consider the i^{th} and j^{th} observations from the k^{th} cluster

- If we let $\mathsf{E}[Y_{ki}] = \mu_{ki}$ and $\mathsf{E}[Y_{kj}] = \mu_{kj}$, then

$$\mathsf{E}[Y_{ki}Y_{kj}] \;\leq\; \min\{\mu_{ki},\; \mu_{kj}\}$$

so that

$$\mathsf{Cov}[Y_{ki}, Y_{kj}] \;\leq\; \min\{\mu_{ki}(1-\mu_{kj}),\; \mu_{kj}(1-\mu_{ki})\}$$

maximum value of p(1-p) = 0.25, is constrained

and

$$\mathsf{Cor}[Y_{ki}, Y_{kj}] \;\leq\; \min\left\{\sqrt{\frac{\mu_{ki}(1-\mu_{kj})}{\mu_{kj}(1-\mu_{ki})}},\; \sqrt{\frac{\mu_{kj}(1-\mu_{ki})}{\mu_{ki}(1-\mu_{kj})}}\right\}$$

- For example, if $\mu_{ki} = 0.3$ and $\mu_{kj} = 0.1$, then Cor$[Y_{ki}, Y_{kj}] \leq 0.26$

- Arguably, a more 'natural' measure of dependence between a pair of binary responses is the odds ratio:

$$\theta_{k,ij} \;=\; \frac{\Pr(Y_{ki} = 1, Y_{kj} = 1)\,\Pr(Y_{ki} = 0, Y_{kj} = 0)}{\Pr(Y_{ki} = 1, Y_{kj} = 0)\,\Pr(Y_{ki} = 1, Y_{kj} = 0)}$$

  - ⋆ not constrained by the marginal means since log $\theta_{k,ij} = \in (-\infty,\ \infty)$
  - ⋆ has a simple interpretation

- If we let $\pi_{k,ij} = \Pr(Y_{ki} = 1, Y_{kj} = 1)$, it is straightforward to show that:

$$\theta_{k,ij} \;=\; \frac{\pi_{k,ij}(1 - \mu_{ki} - \mu_{kj} + \pi_{k,ij})}{(\mu_{ki} - \pi_{k,ij})(\mu_{kj} - \pi_{k,ij})}$$

  - ⋆ re-arrange this to see that $\pi_{k,ij}$ (which is equal to E$[Y_{ki}Y_{kj}]$) is uniquely determined by (μ_{ki}, μ_{kj}, $\theta_{k,ij}$), so that the induced correlation, $\rho_{k,ij}$, is also uniquely determined

- We can therefore, as before, embed this approach within the same two sets of estimating equations:

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1}(\boldsymbol{Y}_k - \boldsymbol{\mu}_k) \\ \mathbf{0} &= \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1}(\boldsymbol{Z}_k - \boldsymbol{\rho}_k) \end{aligned}$$

  - ⋆ key difference is in the model specification for the components of $\boldsymbol{\rho}_k$
  - ⋆ decisions regarding $\boldsymbol{W}_k$ don't really change
  - ⋆ functional form of $\boldsymbol{E}_k$ will be different

- Within this framework, Lipsitz, Laird and Harrington (1991) proposed to analyze repeated measures binary data by using the paired models:

$$\begin{aligned} \text{logit}(\mu_{ki}) &= \boldsymbol{X}_{ki}\boldsymbol{\beta} \\ \log(\theta_{k,ij}) &= \widetilde{\boldsymbol{X}}_{k,ij}\boldsymbol{\alpha} \end{aligned}$$

Modeling the variance-covariance matrix directly

- Finally, rather than augmenting the estimating equations for the mean of $\boldsymbol{Y}_k$ with estimating equations for all pairwise correlations, some folks specify a second set of estimating equations for the variances and covariances:

$$\boldsymbol{U}_K(\boldsymbol{\alpha};\boldsymbol{\beta}) = \sum_{k=1}^{K} \{\boldsymbol{E}_k^\dagger\}^T \{\boldsymbol{W}_k^\dagger\}^{-1} (\boldsymbol{Z}_k^\dagger - \boldsymbol{\sigma}_k) = \mathbf{0}$$

  ⋆ $\boldsymbol{\sigma}_k = (\sigma_{k,11},\ \sigma_{k,12},\ \ldots,\ \sigma_{k,22},\ \ldots)$ is a vector of length $n_k^\dagger = n_k + n_k(n_k-1)/2$ with elements taken from the adopted working covariance matrix:

$$\boldsymbol{V}_k = \begin{bmatrix} \sigma_{k,11} & \sigma_{k,12} & \cdots & \sigma_{k,1n_k} \\ \sigma_{k,21} & \sigma_{k,22} & \cdots & \sigma_{k,2n_k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k,n_k1} & \sigma_{k,n_k2} & \cdots & \sigma_{k,n_kn_k} \end{bmatrix}$$

- $\star$ $\boldsymbol{Z}_k^\dagger = (Z_{k,11}^\dagger,\ Z_{k,12}^\dagger,\ \ldots,\ Z_{k,22}^\dagger,\ \ldots)$ is a vector of length $n_k^\dagger$ with elements:

$$Z_{k,ij}^\dagger \;=\; (Y_{ki} - \mu_{ki})(Y_{kj} - \mu_{kj})$$

- $\star$ $\boldsymbol{W}_k^\dagger$ is an $n_k^\dagger \times n_k^\dagger$ working covariance matrix for $\boldsymbol{Z}_k^\dagger$
- $\star$ $\boldsymbol{E}_k^\dagger = \partial \boldsymbol{\sigma}_k / \partial \boldsymbol{\alpha}$ is a $n_k^\dagger \times q$ matrix where q is the length of $\boldsymbol{\alpha}$

- As with previous formulations, one could consider building paired models for the mean and covariances parameters:

$$\begin{aligned} g_\mu(\mu_{ki}) &= \boldsymbol{X}_{ki}\boldsymbol{\beta} \\ g_\sigma(\sigma_{k,ij}) &= \widetilde{\boldsymbol{X}}_{k,ij}\boldsymbol{\alpha} \end{aligned}$$

- $\star$ $\widetilde{\boldsymbol{X}}_{k,ij}$ is a vector of covariates that are functions of the covariates for the i^{th} and j^{th} study units in the k^{th} cluster
- $\star$ $g_\sigma(\cdot)$ is a user-chosen link function

GEE 2.0

- GEE 1.0 proceeds by iterating between:

$$(1) \quad \widehat{\boldsymbol{\beta}} | \, \widehat{\boldsymbol{\alpha}} : \quad \mathbf{0} \; = \; \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k)$$

$$\text{and} \quad (2) \quad \widehat{\boldsymbol{\alpha}} | \, \widehat{\boldsymbol{\beta}} : \quad \text{moment-based estimation}$$

- GEE 1.5 proceeds by iterating between:

$$(1) \quad \widehat{\boldsymbol{\beta}} | \, \widehat{\boldsymbol{\alpha}} : \quad \mathbf{0} \; = \; \sum_{k=1}^{K} \boldsymbol{D}_k^T \boldsymbol{V}_k^{-1} (\boldsymbol{Y}_k - \boldsymbol{\mu}_k)$$

$$\text{and} \quad (2) \quad \widehat{\boldsymbol{\alpha}} | \, \widehat{\boldsymbol{\beta}} : \quad \mathbf{0} \; = \; \sum_{k=1}^{K} \boldsymbol{E}_k^T \boldsymbol{W}_k^{-1} (\boldsymbol{Z}_k - \boldsymbol{\rho}_k)$$

$$\text{or} \quad (2) \quad \widehat{\boldsymbol{\alpha}} | \, \widehat{\boldsymbol{\beta}} : \quad \mathbf{0} \; = \; \sum_{k=1}^{K} \{\boldsymbol{E}_k^\dagger\}^T \{\boldsymbol{W}_k^\dagger\}^{-1} (\boldsymbol{Z}_k^\dagger - \boldsymbol{\sigma}_k)$$

- In GEE 2.0, estimates of $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$ are obtained via a *joint estimating equation*
  - ⋆ Prentice and Zhao (1991)

- For example, consider the joint estimating equation for the marginal mean parameters and the parameters for the variance-covariance matrix:

$$\boldsymbol{U}_K(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \sum_{k=1}^{K} \begin{bmatrix} \partial \boldsymbol{\mu}_k / \partial \boldsymbol{\beta} & \partial \boldsymbol{\sigma}_k / \partial \boldsymbol{\beta} \\ \partial \boldsymbol{\mu}_k / \partial \boldsymbol{\alpha} & \partial \boldsymbol{\sigma}_k / \partial \boldsymbol{\alpha} \end{bmatrix}^T \begin{bmatrix} \boldsymbol{V}_k & \boldsymbol{C}_k \\ \boldsymbol{C}_k^T & \boldsymbol{W}_k^\dagger \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Y}_k - \boldsymbol{\mu}_k \\ \boldsymbol{Z}_k^\dagger - \boldsymbol{\sigma}_k \end{bmatrix}$$

  - ⋆ $\boldsymbol{V}_k$ is the working model for Cov$[\boldsymbol{Y}_k]$
  - ⋆ $\boldsymbol{W}_k^\dagger$ is the working model for Cov$[\boldsymbol{Z}_k]$
  - ⋆ $\boldsymbol{C}_k$ is the working model for Cov$[\boldsymbol{Y}_k, \boldsymbol{Z}_k]$

- Involves complete specification of all third and fourth moments
  - ⋆ first and second are insufficient to specify $\boldsymbol{W}_k^\dagger$ and $\boldsymbol{C}_k$

- As in GEE 1.0 and GEE 1.5, one has substantial choice in the specification of $\boldsymbol{C}_k$
  - ⋆ a simple choice is to take $\boldsymbol{C}_k = \mathbf{0}$
  - ⋆ in principle, other choices will lead to more efficient estimation of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$
- Note, specification of all moments leads to maximum likelihood
  - ⋆ hence, the more moments one (correctly) specifies the more efficient the estimator will be
- Regardless of the specific choice, one can derive the form of the variance-covariance matrix of the asymptotic sampling distribution of the estimator $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\alpha}})$
  - ⋆ again of 'sandwich' form
  - ⋆ robust to misspecification of $\boldsymbol{V}_k$, $\boldsymbol{C}_k$, and $\boldsymbol{W}_k^{\dagger}$

Summary

- Focus on estimation/inference for *marginal* models
  - ⋆ $\boldsymbol{\beta}$ are the regression coefficients of interest
  - ⋆ $\boldsymbol{\alpha}$ are (working) covariance parameters, viewed primarily as a nuisance

- Three frameworks that vary in how $\boldsymbol{\alpha}$ is specified and estimated

  GEE 1.0: moment-based estimation of $\boldsymbol{\alpha}$

  GEE 1.5: simultaneous model-based estimation of $\boldsymbol{\alpha}$

  GEE 2.0: joint model-based estimation of $\boldsymbol{\alpha}$

- A by-product of the use of additional estimating equations in GEE 1.5 and GEE 2.0 is that one can obtain measures of uncertainty for estimates of $\boldsymbol{\alpha}$

- For all three:
  - ⋆ if the mean model is correctly specified then the point estimates for $\boldsymbol{\beta}$ will be consistent and asymptotically Normal
  - ⋆ the 'robust' sandwich form of the variance yields valid standard errors, regardless of whether the choice of the working covariance structure is correct (in large samples, at least)
    - ∗ 95% CI's have the appropriate coverage rates
    - ∗ hypothesis testing based on the Wald statistic is valid

- Diggle, Heagerty, Liang and Zeger (2002)

  When regression coefficients are the scientific focus as in the examples here, one should invest the lion's share of time in modeling the mean structure, while using a reasonable approximation to the covariance. The robustness of the inferences about can be checked by fitting a final model using different covariance assumptions and comparing the two sets of estimates and their robust standard errors. If they differ substantially, a more careful treatment of the covariance model may be necessary.