Advanced Methods in Biostatistics II

Lecture 11

November 30, 2017

Correlated data

- Non-independence of observations may result from either serial correlation or clustering of the observations.
- Serial correlation is present when observations are measured at equally spaced time points and the error terms from adjacent time points are correlated.
- Cluster correlation is present when the observations exhibit a group structure, and observations within the group are correlated with one another.

Examples

- Longitudinal data - repeated observations on each subject measured at different time points.
- Clustered data - subjects are grouped in some way (i.e., several members from the same family).
- Multilevel data - multiple levels of groupings (i.e., students nested within classrooms nested within schools).

Mixed models

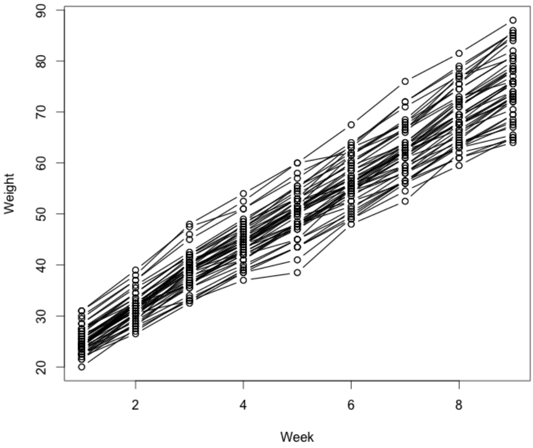
- One approach for dealing with such data is to think of certain parameters as being random rather than fixed.
- This gives rise to so-called mixed models.
- This is particularly useful when the natural asymptotics of the problem have the number of parameters tending to infinity with the sample size.

Example

- Let us illustrate with the `pig.weights` dataset which is available from the `SemiPar` package.
- It consists of 9 repeated weight measures on 48 pigs.
- This data frame contains the following columns:

```
%id.num      identification number of pig.
%num.weeks   number of weeks since measurements commenced.
%weight      bodyweight of pig "id.num" after "num.weeks" weeks.
```

Example



Example

- Let us consider a model of the form

$$y_{ij} = \mu_i + \beta_1 x_{ij} + \epsilon_{ij},$$

  where $i = 1, \ldots 48$ represents the pig and $j = 1, \ldots 9$ represents the week.

- As stated, this is an example of a standard analysis of covariance (ANCOVA) model.

Example

- Note, if we increase the number of pigs included in the study, we need to similarly increase the number of parameters.
- In addition, results will be highly dependent on the pigs included in the study, and may not generalize to the larger population of pigs.

Random intercept model

- Consider replacing μ_i with a random intercept term.
- The model can then be written:

$$y_{ij} = \beta_0 + u_i + \beta_1 x_{ij} + \epsilon_{ij}.$$

  for $i = 1, \ldots m$, $j = 1, \ldots n_i$, where $u_i \sim_{iid} N(0, \sigma_u^2)$ is treated as a random sample from a distribution with $\sigma_u^2 > 0$.
- Here u_i and ϵ_{ij} are assumed to be independent and $N = \sum n_i$.

Random effects

- Here the term u_i is referred to as a random effect.
- Importantly, the collection of u_i's can be described by a single parameter σ_u^2, which is called a variance component.
- This allows us to incorporate the randomness that would occur if we were to take another sample of pigs.

Mixed model

- This is an example of a mixed model, or mixed-effects model.
- It consists of fixed effects:

$$\beta_0 + \beta_1 x_{ij},$$

  and a random effect:

$$u_i \sim_{iid} N(0, \sigma_u^2).$$

Mixed model

- The main advantage of using mixed models over standard fixed effects models is that they can help avoid overfitting.
- We only estimate the variance of the random effects as opposed to the random effects themselves.
- In addition, there are more degrees of freedom available to estimate the fixed effects.

Interpretation

- It is useful to think of the distribution of the random effect as a population distribution (i.e., the population of pigs in our example).
- We can write the model as follows:

$$\begin{aligned} y_{ij} &= \alpha_i + \beta_1 x_{ij} + \epsilon_{ij} \\ \alpha_i &= \beta_0 + u_i \end{aligned}$$

where $u_i \sim N(0, \sigma_u^2)$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ and $\mathrm{cov}(u_i, \epsilon_{ij}) = 0$.

Interpretation

- Another way to think about random effects is to consider a fixed effect treatment of the u_i terms.
- Since we included an intercept, we need to add a linear constraint on the u_i for identifiability, i.e.,

$$\sum_{i=1}^{m} u_i = 0.$$

- β_0 is interpreted as the overall mean and the u_i terms are the pig-specific deviation around that mean.
- The random effect model simply specifies that the u_i are iid $N(0, \sigma_u^2)$ and mutually independent from ϵ_{ij}.

Random intercept model

- The inclusion of the random intercept allows for the modeling of the within-group correlation.
- Note that

$$
\begin{aligned}
\mathrm{cov}(y_{ij}, y_{i'j'}) &= \mathrm{var}(u_i + \epsilon_{ij}, u_{i'} + \epsilon_{i'j'}) \\
&= \mathrm{cov}(u_i, u_{i'}) + \mathrm{cov}(\epsilon_{ij}, \epsilon_{i'j'}) \\
&= \begin{cases} \sigma^2 + \sigma_u^2 & \text{if } i = i' \text{ and } j = j' \\ \sigma^2 & \text{if } i = i' \text{ and } j \neq j' \\ 0 & \text{otherwise} \end{cases}
\end{aligned}
$$

Random intercept model

- Thus the correlation between observations within the same group is given by:

$$\frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}.$$

- This is the ratio between the between-group variability, σ_u^2, and the total variability, $\sigma_u^2 + \sigma^2$.
- This coefficient is often called the intra-class correlation coefficient, and used in the context of reliability.

Variance components model

- Because the variance of the observations has been partitioned into two components these models are sometimes called variance component models.
- Here σ_u^2 represents between-group variability and σ^2 represents the within-group variability.

General form

- The general linear mixed model can be written as follows:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon.$$

where $E(\mathbf{u}) = E(\varepsilon) = \mathbf{0}$, $\mathrm{var}(\mathbf{u}) = \Sigma_u$, $\mathrm{var}(\varepsilon) = \Sigma_\epsilon$, and $\mathrm{cov}(\mathbf{u}, \varepsilon) = \mathbf{0}$.

- Here, $\mathbf{Z}$ is a matrix of fixed predictors, used to specify membership in the various clusters or subgroups.

Example

- For the pig data set:

$$
\begin{aligned}
\mathbf{y} &= \begin{pmatrix} y_{1,1} & \dots & y_{1,9} & \dots & y_{48,1} & \dots & y_{48,9} \end{pmatrix}' \\
\mathbf{X} &= \begin{pmatrix} 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 1 & \dots & 9 & \dots & 1 & \dots & 9 \end{pmatrix}' \\
\boldsymbol{\beta} &= \begin{pmatrix} \beta_0 & \beta_1 \end{pmatrix}' \\
\mathbf{Z} &= \mathbf{I}_{48} \otimes \mathbf{J}_9 \qquad \mathbf{u} = \begin{pmatrix} u_1 & \dots & u_{48} \end{pmatrix}' \\
\Sigma_u &= \sigma_u^2 \mathbf{I} \qquad \Sigma_\epsilon = \sigma_u^2 \mathbf{I}
\end{aligned}
$$

Estimation

- We can derive an estimate of β by re-writing the linear mixed model as follows:

  $$\mathbf{y} = \mathbf{X}\beta + \varepsilon^*$$

  where $\varepsilon^* = \mathbf{Zu} + \varepsilon$.

- This is a standard linear model with correlated errors, where

  $$\text{var}(\varepsilon^*) \equiv \mathbf{V} = \mathbf{Z}\Sigma_u\mathbf{Z}' + \Sigma_\epsilon.$$

Estimation

- For a given $\mathbf{V}$ we can estimate β using generalized least-squares:

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}.$$

- This particular estimator is the best linear unbiased estimator (BLUE).
- If $\mathbf{y}$ follows a multivariate normal distribution it is also the maximum likelihood estimator (MLE) and the uniformly minimum variance unbiased estimator (UMVUE).

Prediction

- In mixed models it is also possible to predict the unobserved vector $\mathbf{u}$.
- Because $\mathbf{u}$ is a random vector rather than a fixed parameter, we predict rather than estimate $\mathbf{u}$.
- This ability is one of the key attributes of mixed models.
- We seek a Best Linear Unbiased Predictor (BLUP) for $\mathbf{u}$, which we denote $\hat{\mathbf{u}}$.

Prediction

- The term $\hat{\mathbf{u}}$ is the BLUP if:
    1. It is a linear function of $\mathbf{y}$;
    2. It is unbiased;
    3. $E[(\hat{\mathbf{u}} - \mathbf{u})^2] \leq E[(\mathbf{v} - \mathbf{u})^2]$ for any other linear and unbiased predictor $\mathbf{v}$.
- The BLUP for $\mathbf{u}$ is given by $E(\mathbf{u}|\mathbf{y})$.

Prediction

- Consider the univariate case: $E[(u-\theta(y))^2]$

$$
\begin{aligned}
& E[(u-\theta(y))^2] \\
& \quad = E[(u-E[u|y]+E[u|y]-\theta(y))^2] \\
& \quad = E[(u-E[u|y])^2]-2E[(u-E[u|y])(E[u|y]-\theta(y))] \\
& \qquad + E[(E[u|y]-\theta(y))^2] \\
& \quad = E[(u-E[u|y])^2]+E[(E[u|y]-\theta(y))^2] \\
& \quad \geq E[(u-E[u|y])^2].
\end{aligned}
$$

- Note $E(E[(u-E[u|y])(E[u|y]-\theta(y))]|y)=0$
- Hence, $\theta(y)=E[u|y]$.

Prediction

- Note this is the best predictor, regardless of the setting.
- In the context of linear models, this predictor is both linear (in $\mathbf{y}$) and unbiased in the sense that:

$$E[u - E[u|y]] = 0.$$

- Therefore, even in the more restricted class of linear unbiased estimators, $E[u|y]$ remains best.

Review - Conditional distributions

- Suppose that $\mathbf{y}_1$ and $\mathbf{y}_2$ are jointly multivariate normal with $\Sigma_{12} \neq 0$, i.e.

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right).$$

- The conditional distribution of $\mathbf{y}_2 \mid \mathbf{y}_1$ is $N(\boldsymbol{\mu}_{\mathbf{y}_2|\mathbf{y}_1}, \Sigma_{\mathbf{y}_2|\mathbf{y}_1})$, where

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{y}_2|\mathbf{y}_1} &= \boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1) \\ \Sigma_{\mathbf{y}_2|\mathbf{y}_1} &= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}'. \end{aligned}$$

Prediction

- Since **u** and **y** are jointly multivariate normal, we have

$$E(\mathbf{u}|\mathbf{y}) = E(\mathbf{u}) + \mathrm{cov}(\mathbf{u}, \mathbf{y})(\mathrm{var}(\mathbf{y}))^{-1}(\mathbf{y} - E(\mathbf{y})).$$

- Note:

$$\begin{aligned} \mathrm{var}(\mathbf{y}) &= \mathbf{V} \\ \mathrm{cov}(\mathbf{u}, \mathbf{y}) &= \mathrm{cov}(\mathbf{u}, \mathbf{Zu}) = \Sigma_u \mathbf{Z}'. \end{aligned}$$

- Hence:

$$E(\mathbf{u}|\mathbf{y}) = E(\mathbf{u}) + \Sigma_u \mathbf{Z}' \mathbf{V}^{-1}(\mathbf{y} - E(\mathbf{y})).$$

Prediction

- A complication arises in that we typically do not know the variance components.
- As that is the case, we must plug in estimates (obtained either using REML or ML).
- The BLUPs then lose their optimality properties and are often called EBLUPs (for empirical BLUPs).

Example

- Let us revisit the pig example.
- We can predict u_i by considering the estimate $E[u_i \mid \mathbf{y}]$.
- To derive this, note the density for $u_i \mid \mathbf{y}$ is equal to the density of $u_i \mid \mathbf{y}_i$, since u_i is independent of every $y_{i'j}$ for $i \neq i'$.
- Hence, it is sufficient to consider $E(u_i|\mathbf{y}_i)$.

Example

- Here

$$\begin{aligned}\mathrm{var}(\mathbf{y}_i) &= \mathbf{V}_i \\ &= \mathbf{J}_{n_i}\Sigma_u\mathbf{J}'_{n_i} + \Sigma_\epsilon \\ &= \sigma_u^2\mathbf{J}_{n_i}\mathbf{J}'_{n_i} + \sigma_\epsilon^2\mathbf{I}.\end{aligned}$$

and

$$\begin{aligned}\mathrm{cov}(u_i, \mathbf{y}) &= \mathrm{cov}(u_i, \mathbf{J}_{n_i}u_i) \\ &= \Sigma_u\mathbf{J}'_{n_i} \\ &= \sigma_u^2\mathbf{J}'_{n_i}\end{aligned}$$

Matrix result

Lemma

$$(a\mathbf{I} + b\,\mathbf{J}_n\mathbf{J}_n')^{-1} = \frac{1}{a}\left(\mathbf{I} - \frac{b}{a+nb}\mathbf{J}_n\mathbf{J}_n'\right)$$

for $a \neq 0$ and $a \neq -nb$.

Example

- Using the lemma, we can write:

$$\hat{u}_i = \frac{n_i\sigma_u^2}{\sigma^2 + n_i\sigma_u^2}(\bar{\mathbf{y}}_i - \beta_0 - \beta_1\bar{\mathbf{x}}_i).$$

  where $\bar{\mathbf{y}}_i$ is the average weight of the i^{th} pig and $\bar{\mathbf{x}}_i$ is the average week value.

- This is a type of shrinkage, where the mean residual for the i^{th} pig is shrunk toward 0 with a shrinkage factor given by

$$\frac{n_i\sigma_u^2}{\sigma^2 + n_i\sigma_u^2}.$$

Example

- The larger the between-group variance σ_u^2 is relative to the within-group variance σ^2, the less shrinkage we have.
- In addition, the more observations per group (i.e., pig), the less shrinkage we have.
- In this way prediction is calibrated to weigh the contribution of the individual pig versus the contribution of the others.