

Longitudinal Data: Repeated Measures

Linear Models with repeated measures

Alan Hubbard

Oct 2, 2018

Introduction

- ▶ In last section, we concentrated on ways to reduce repeated, longitudinal outcome data into a single summary measure (e.g., count of binary repeated measures) so that standard regression procedures could be used.
- ▶ However, in this section, we explore how linear regression can accommodate repeated measures.
- ▶ This will lead ultimately to describing popular repeated measures procedures, like Mixed Models and Generalized Estimating Equations (GEE).
- ▶ First, we examine data from an observational study on the size of kids mouths as they grew.

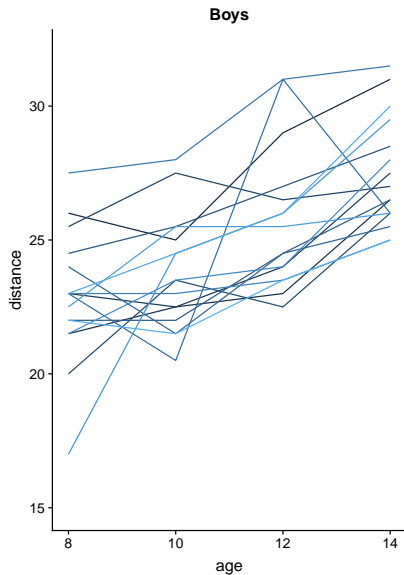
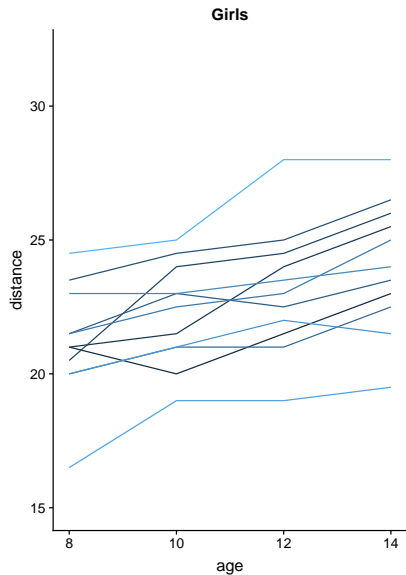
```
dat <- read_csv("dental.csv")
tbl_df(dat)
```

```
## # A tibble: 108 x 5
##   obsno child  age distance gender
##   <int> <int> <int>     <dbl>   <int>
## 1      1      1      8      21       0
## 2      2      1     10      20       0
## 3      3      1     12     21.5       0
## 4      4      1     14      23       0
## 5      5      2      8      21       0
## 6      6      2     10     21.5       0
## 7      7      2     12      24       0
## 8      8      2     14     25.5       0
## 9      9      3      8     20.5       0
## 10     10      3     10      24       0
## # ... with 98 more rows
```

Data Notation

- ▶ $\text{child} = i$ (id)
- ▶ $\text{age} = X_{ij1}$
- ▶ $\text{gender} = X_{ij2}$: 1 is male, 0 female
- ▶ measure of teeth $\text{distance} = Y_{ij}$

Plot the data



Plots show:

1. variation among outcome at start (age 8 years),
2. increase over time, but not perfectly linear for subjects,
3. maybe one measurement error (boy that drops from age 12 to 14).

Look at basic summaries with dplyr

```
# get number of obs per person and mean  
sum_dat=group_by(dat,child) %>% summarise(mean=mean(distance))  
table(sum_dat$gender)
```

```
##
```

```
##  0  1
```

```
## 11 16
```

Linear model of distance by age and gender

- ▶ Y_{ij} distance of the j th measurement on the i th child,
- ▶ X_{ij1} is the age of i th child, j th measurement,
- ▶ X_{ij2} is the gender of i th child, j th measurement (1 is male, 0 is female) - same for all $j = 1, \dots, 4$ measurements.

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2} + e_{ij}, E(e_{ij} | X_{ij}) = 0$$

Ordinary Least Squares

- ▶ The name is based on the objective function, that is:

$$\hat{\beta} \equiv \operatorname{argmin}_{\beta} \sum_{i=1}^n \sum_{j=1}^4 (Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}])^2$$

- ▶ So, it minimizes the mean-squared error, or the squared residuals $r_{ij} \equiv Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]$
- ▶ Can be motivated in two ways
 - ▶ Maximum likelihood estimation by assuming, for example, $e_{ij} \sim N(0, \sigma_e)$
 - ▶ As the procedures that minimizes the squared residuals, regardless of distribution of e_{ij}

MLE Justification

- ▶ Assume independent and identically distributed $e_{ij} \sim N(0, \sigma_e)$ (so for now, observations are independent regardless if they are made on same child).
- ▶ The likelihood is defined as simply the probability of the observed data. In this case, we can define the likelihood of one observation as: $P(Y_{ij} | X_{ij1}, X_{ij2})P(X_{ij1}, X_{ij2})$
- ▶ However, we can ignore the 2nd term (our parameters of interest are not a function of the distribution of $P(X_{ij1}, X_{ij2})$).

MLE , cont.

- ▶ The likelihood of all the observations is (if they are all independent) the probability of each observation multiplied together (e.g., $P(A \& B) = P(A)P(B)$ if independent).

$$\begin{aligned} &= P(Y_{11} \mid X_{111}, X_{112})P(Y_{12} \mid X_{121}, X_{122}) \\ &\cdots P(Y_{m3} \mid X_{m31}, X_{m32})P(Y_{m4} \mid X_{m41}, X_{m42}) \\ &= \prod_{i=1}^m \prod_{j=1}^4 P(Y_{ij} \mid X_{ij1}, X_{ij2}) \end{aligned}$$

- ▶ If we assume a normal distribution for the e_{ij} , then the distribution of $Y_{ij} \mid X_{ij1}, X_{ij2}$ is also normal, since the rest of the linear equation is fixed when the X's are fixed.

Distribution of $Y|X$ in this model is distribution of e .

- Note that $e_{ij} = Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]$, so that

$$\begin{aligned} P(Y_{ij} \mid X_{ij1}, X_{ij2}) &= P(e_{ij} \mid X_{ij1}, X_{ij2}) \\ &= P(Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]) \\ &= \phi\left(\frac{Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]}{\sigma_e}\right) \end{aligned}$$

, where $\phi(x)$ represents the standard normal density $= \frac{\exp(-x^2/2)}{\sqrt{2\pi}}$.

Likelihood function

- ▶ Thus, for all the data the likelihood (as a function of β) is

$$L(\beta \mid \text{data}) = \prod_{i=1}^m \prod_{j=1}^4 \phi\left(\frac{Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]}{\sigma_e}\right)$$

- ▶ For several reasons, the log of this function is maximized (the maximum of the function, w.r.t. β or its log is the same)

$$l(\beta \mid \text{data}) = \sum_{i=1}^m \sum_{j=1}^4 \log\left[\phi\left(\frac{Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]}{\sigma_e}\right)\right]$$

- ▶ taking the log of the ϕ function is

$$\begin{aligned} & \log\left(\frac{\exp\left(-\frac{\{Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]\}^2}{2\sigma_e^2}\right)}{\sqrt{2\pi\sigma_e^2}}\right) \\ &= -\left(\frac{\{Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}]\}^2}{2\sigma_e^2}\right) - 1/2 * \log(2\pi\sigma_e^2) \end{aligned}$$

Maximizing likelihood

- ▶ So, maximizing the log-likelihood is the same as minimizing the

$$MSE(\beta) \equiv \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^4 (Y_{ij} - [\beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}])^2$$

.

- ▶ Thus one can define OLS as

$$\hat{\beta} \equiv \operatorname{argmin}_{\beta} MSE(\beta)$$

Properties of MLE

- ▶ If you believe a particular parametric model for the data-generating distn, then: MLE's have **invariance property**:
 - ▶ if $\hat{\beta}_n$ is the MLE, then any function of them is the MLE estimate, $h(\hat{\beta}_n)$ is the MLE of $h(\beta_{true})$.
 - ▶ e.g., if $\hat{\beta}_n$ is MLE, then $e^{\hat{\beta}_n}$ is also an MLE.
- ▶ **Asymptotically unbiased** (consistent), that is for any

$$\epsilon > 0, \lim_{n \rightarrow \infty} P(|\hat{\beta}_n - \beta_{true}| > \epsilon) \rightarrow 0.$$

- ▶ **Asymptotically linear**: $\hat{\beta}_n \approx \beta_{true}$ plus i.i.d. terms with mean 0.
- ▶ **Asymptotically efficient**: if $\tilde{\beta}_n$ is a competing estimator in same model, then

$$\lim_{n \rightarrow \infty} \left[\frac{\text{var}(\hat{\beta}_n)}{\text{var}(\tilde{\beta}_n)} \right] \leq 1$$

Motivating OLS as estimating equation

- ▶ OLS also can be motivated without knowing the likelihood of the data (e.g., without asserting that $e \sim N(0, \sigma_e)$).
- ▶ It turns out that the estimating equation is consistent for β regardless of the distribution of $e \mid X_{ij}$.
- ▶ In general, an estimating equation results in a consistent estimator of β_{true} if $E_{\beta_{true}} g(O_i, \beta_{true}) = 0$, where $g(\cdot)$ is an estimating equation, and O_i is the data on one statistical unit, so in our case, $O_i = (Y_{ij}, \vec{X}_{ij}, j = 1, \dots, 4)$.
- ▶ These leads to an estimator by solving $\frac{1}{m} \sum_{i=1}^m g(O_i, \beta) = 0$.

OLS as Solution to minimizing MSE

- ▶ One has a function that one is trying to minimize w.r.t. β , i.e., $\operatorname{argmin}_{\beta} \operatorname{MSE}(\beta)$
- ▶ Let's take a simple linear model for now with one predictor:
 $Y_{ij} = \beta_0 + \beta_1 X_{ij} + e$, so minimizing the MSE is same as minimizing:
- ▶ Minimizing the function specifically in this case, is minimizing

$$\operatorname{MSE}(\beta) = \sum_{i=1}^n \sum_{j=1}^4 (Y_{ij} - [\beta_0 + \beta_1 X_{ij}])^2$$

.

Minimizing function

- ▶ How does one minimize functions?
 - ▶ Try an exhaustive grid over all possible values of the parameters $\beta = (\beta_0, \beta_1)$ and find the combo that has smallest $MSE(\beta)$
 - ▶ Find the solution analytically using calculus.
- ▶ One can solve a set of two differential equations in this case, so:

$$\frac{\partial MSE(\beta)}{\partial \beta_0} = 0$$

and

$$\frac{\partial MSE(\beta)}{\partial \beta_1} = 0$$

- ▶ Solve for (β_0, β_1) .
- ▶ Show on board.
- ▶ In this case, one gets the following general solution (in matrix form):

$$\hat{\vec{\beta}}_n = (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \vec{Y}$$

Is this a consistent estimator?

- ▶ We now from theory, if it's the MLE, it is a consistent estimator (see above).
- ▶ However, is it more generally consistent than that?
- ▶ Just have to plug and chug to see what is the expected value of this estimator, $E(\hat{\beta}_n)$ and whether it is equal to the true β_{true} .
- ▶ This one is “easy”:

$$\begin{aligned} E(\hat{\beta}_n \mid \mathbf{X}) &= E\{(\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \vec{Y} \mid \mathbf{X}\} \\ &= (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T E(\vec{Y} \mid \mathbf{X}) \\ &= (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}^T \mathbf{X} \vec{\beta}_{true} = \vec{\beta}_{true} \end{aligned}$$

What about impact of repeated measures (dependent data)?

- ▶ The dependence among observations on same subject does not impact the “proof”
 - ▶ This is because expectation of sums is sum of expectations, e.g.,
$$E(\sum_{i=1}^m \sum_{j=1}^{n_i} Y_{ij}) = \sum_{i=1}^m \sum_{j=1}^{n_i} E(Y_{ij})$$
 - ▶ Note, this is NOT generally true for variances of sum - that the variance of sum is only equal to sum of variances if the random variables are independent.
- ▶ Thus, it looks like ignoring the dependence does not necessarily cause bias in estimation of the coefficients (generally true, but gets complicated in more complicated setting with unequal numbers of measurements per unit, and if the linear model is misspecified).

Variance of coefficient estimates from OLS fit ($var(\hat{\beta}_n)$).

- ▶ Note that $\hat{\beta}_n = (\hat{\beta}_{0,n}, \hat{\beta}_{1,n})$ is a vector the variance of a vector implies the variance-covariance matrix, or a matrix with the variance of all parameter estimates and all pairwise covariances.
- ▶ So, when we say we want the $var(\hat{\beta}_n)$, which we sometimes represent as $V_{\hat{\beta}_n}$, it means the estimate of each component of the following in our simple linear model case ($Y_{ij} = \beta_0 + \beta_1 X_{ij} + e$), so

$$V_{\hat{\beta}_n} = \begin{bmatrix} var(\hat{\beta}_{0,n}) & cov(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}) \\ cov(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}) & var(\hat{\beta}_{1,n}) \end{bmatrix}$$

- ▶ We want to derive an estimate of this (since we never know it) so

$$\hat{V}_{\hat{\beta}_n} = \begin{bmatrix} \hat{var}(\hat{\beta}_{0,n}) & \hat{cov}(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}) \\ \hat{cov}(\hat{\beta}_{0,n}, \hat{\beta}_{1,n}) & \hat{var}(\hat{\beta}_{1,n}) \end{bmatrix}$$

Standard errors

- ▶ Note that the standard errors reported in regression outputs are based upon this estimated matrix, e.g.,

$$SE(\hat{\beta}_{1,n}) = \sqrt{\hat{var}(\hat{\beta}_{1,n})}.$$

- ▶ If we ever want to report an estimated parameter that involves more than one coefficient estimate, then we also need the covariance terms (though we can have procedures that automate this process).

Review some rules of expectation and variance for random variables and vectors.

- Expectation of sum is sum of expectations

$$E\left(\sum_{i=1}^m \sum_{j=1}^{n_i} Y_{ij}\right) = \sum_{i=1}^m \sum_{j=1}^{n_i} E(Y_{ij})$$

- Expectation of constant times random variable (vector) is constant times expectation of random variable (vector)

$$E(AY) = AE(Y)$$

$$E(\vec{A}^T \vec{Y}) = \vec{A}^T E(\vec{Y})$$

- Variance of constant times random variable is constant squared times variance of random variable

$$\text{var}(AY) = A^2 \text{var}(Y)$$

$$\text{var}(\vec{A}^T \vec{Y}) = \vec{A}^T V_{\vec{Y}} \vec{A}$$

Using rules to get $\text{var}(\hat{\beta}_n)$

- Get

$$\text{var}(\hat{\beta}_n \mid \mathbf{X}) = \text{var}\{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{Y} \mid \mathbf{X}\}$$

- Not that this is same situation as the matrix version of the variance of a constant times a random variable
- In this case, matrix of constants $((\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)$ times a random vector (\vec{Y}) .
- Using the rule above, this means that:

$$\text{var}\{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \vec{Y} \mid \mathbf{X}\} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T V_{\vec{Y}} [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]^T$$

- Using rules of matrix transformation, one gets:

$$\text{var}(\hat{\beta}_n \mid \mathbf{X}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T V_{\vec{Y}} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$$

- The term “sandwich estimator” of the variance comes from this form where the sandwich filling is the variance-covariance matrix of $\vec{Y} \mid \mathbf{X}$, or $V_{\vec{Y}}$ and the bread is the $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$.

Estimating $\text{var}(\hat{\beta}_n \mid \mathbf{X})$

- ▶ All the matrices \mathbf{X} are known (we are conditioning on them)
- ▶ The only unknown matrix is $V_{\tilde{y}}$, or the variance covariance of the outcomes.
- ▶ Thus, we have to estimate it in order to fill in the unknown components of it.
- ▶ This is where assumptions about the “form” of the variance-covariance (VC) matrix, that is assumptions of how the observations are potentially dependent, have an impact.

a

$Var(\hat{\beta})$ from variance assumptions of OLS
(Independence and constant variance)

$$V(\vec{Y}) =_{N \times N} \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \dots & 0 \\ 0 & \sigma^2 & 0 & 0 \dots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & 0 \dots & \sigma^2 \end{bmatrix}$$

- Inference returned by standard OLS assumes, the observations are assumed to be independent and *homoskedastic* (constant variance).

Estimating $Var(\vec{Y})$
 to get estimate of $Var(\hat{\beta})$
 under assumptions of OLS

- To estimate σ^2 , which is now the variance of the Y_{ij} 's. Note that:

$$\text{var}(Y_{ij} | \mathbf{X}_{ij}) = \text{var}\{e_{ij} | \mathbf{X}_{ij}\} = \sigma^2$$

- We don't know the e_{ij} 's, but we can estimate them from the residuals:

$$r_{ij} = Y_{ij} - \hat{Y}_{ij}(\vec{X}_{ij}) = Y_{ij} - \left\{ \hat{\beta}_0 + \hat{\beta}_1 X_{ij1} + \dots + \hat{\beta}_p X_{ijp} \right\}$$

OLS estimate of $Var(\hat{\beta})$

- To get a estimate of σ^2 , get an average (sort of – adjusted denominator, $N-p$, reduces small sample bias) of the squared residuals:

$$\hat{\sigma}^2 = \frac{1}{N-p} \sum_{i=1}^m \sum_{j=1}^{n_i} (r_{ij})^2$$

- Then, this is plugged back in V_i to get:

$$\hat{V}_i = \hat{\sigma}^2 I_{n_i \times n_i} \text{ or } \hat{V} = \hat{\sigma}^2 I_{N \times N},$$

where I represents identify matrix

$$I_{4 \times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

OLS estimate of $Var(\hat{\beta})$

- Finally, plug this back into to get the estimated variance of the coefficients

$$\begin{aligned}
 \hat{\text{var}}(\hat{\beta} | X) &= (X^T X)^{-1} X^T \hat{V} X (X^T X)^{-1} \\
 &= (X^T X)^{-1} X^T \hat{\sigma}^2 I X (X^T X)^{-1} \\
 &= \hat{\sigma}^2 (X^T X)^{-1} X^T X (X^T X)^{-1} \quad \text{var}(\hat{\beta})_{(p+1) \times (p+1)} = \begin{bmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \dots & \text{cov}(\hat{\beta}_0, \hat{\beta}_p) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) & \text{cov}(\hat{\beta}_1, \hat{\beta}_2) & \dots \\ \dots & \dots & \dots & \dots \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_p) & \text{cov}(\hat{\beta}_1, \hat{\beta}_p) & \dots & \text{var}(\hat{\beta}_p) \end{bmatrix} \\
 &= \hat{\sigma}^2 (X^T X)^{-1}
 \end{aligned}$$

- This is a matrix where the diagonal elements are the estimated variances of the coefficient estimates. The $SE(\hat{\beta})$ are the square-roots of these estimated variances, e.g., $SE(\hat{\beta}_0) = \sqrt{\hat{\text{var}}(\hat{\beta}_0)}$

What if OLS model for is wrong ?

- Ignoring the correlation, as in the standard errors returned by OLS, can give biased estimates of $\text{var}(\hat{\beta})$ (and thus biased $SE(\hat{\beta})$).
- This can lead to erroneous confidence intervals and tests.
- However, we can still use OLS if we can just repair the estimates of $\text{var}(\hat{\beta})$.
- That means getting a better (maybe even) unbiased estimate of $\text{Var}(\hat{Y})$, so choose *bigger variance-covariance model*.

Estimating V under bigger model

- First, we will take the case that measurements are time-structured and the same for each person ($n_i=n$) - the dental example.
- For this type of data, one can assume the most general structure about V (unstructured).
- The goal is to estimate the individual components of V_i (the v_{ijk}) using the residuals of OLS.

Form of the each unit's Variance-Covariance matrix, V_i .

- In order to estimate the V , we will assume that the individual covariance matrices (the V_i) are all the same: $V_1 = V_2 = \dots = V_m$.
- That is equivalent to saying the $v_{ijk} = v_{jk}$, or the variance and covariances of equivalent observations are the same for every individual.
- That way, we will only have to estimate one V_i of in order to build V .
- Thus, we can call each $V_1, V_2, \dots, V_m = V_0$.

The Variance-Covariance of Y

$$Var(\vec{Y}) = \begin{bmatrix} V_0 & 0 & 0 & 0 & \dots & 0 \\ 0 & V_0 & 0 & 0 & \dots & 0 \\ 0 & 0 & V_0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & V_0 \end{bmatrix}$$

Each one of these
represents a matrix,
e.g.

$$V_0(\vec{Y}_1)_{n \times n} = \begin{bmatrix} \text{var}(Y_{11}) & \text{cov}(Y_{11}, Y_{12}) & \dots & \text{cov}(Y_{11}, Y_{1n}) \\ \text{cov}(Y_{11}, Y_{12}) & \text{var}(Y_{12}) & \text{cov}(Y_{12}, Y_{13}) & \dots \\ \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & \text{var}(Y_{1n}) \end{bmatrix}$$

Estimating the components of V_0 .

- If we observed the errors, then it would be easy to estimate the v_{jk} , because, given the X_{ij} 's are treated as fixed, $v_{jk} = \text{cov}(Y_{ij}, Y_{ik}) = \text{cov}(e_{ij}, e_{ik}) = E(e_{ij}e_{ik}) - E(e_{ij})E(e_{ik}) = E(e_{ij}e_{ik})$.
- So estimate cov with average:

$$\hat{v}_{jk} = \frac{1}{m} \sum_{i=1}^m e_{ij}e_{ik}.$$

- We don't observe the errors, but we can still estimate the covariance of any two measurements on the same person using the residuals.

$$\hat{v}_{jk} = \frac{1}{m} \sum_{i=1}^m r_{ij}r_{ik}, r_{ij} = Y_{ij} - \hat{Y}_{ij}(\bar{X}_{ij}), \text{ where } \hat{Y}_{ij}(\bar{X}_{ij}) \text{ is the predicted value of outcome}$$

Plugging Into the Big $V(\vec{Y})$

- Now that one has derived the estimates of each unique variance and covariance among the set of measurements on each unit, then just plug these into to get:

$$\hat{V}_0(\vec{Y})_{n \times n} = \begin{bmatrix} \tilde{ave}(r_{i1}^2) & \tilde{ave}(r_{i1} * r_{i2}) & \dots & \tilde{ave}(r_{i1} * r_{in}) \\ \tilde{ave}(r_{i1} * r_{i2}) & \tilde{ave}(r_{i2}^2) & \dots & \tilde{ave}(r_{i2} * r_{in}) \\ \dots & \dots & \dots & \dots \\ \tilde{ave}(r_{i1} * r_{in}) & \tilde{ave}(r_{i2} * r_{in}) & \dots & \tilde{ave}(r_{in}^2) \end{bmatrix}, \text{ where } \tilde{ave} \text{ represents adusted for small sample}$$

$$\hat{V}(\vec{Y})_{N \times N} = \begin{bmatrix} \hat{V}_{0,n \times n} & 0_{n \times n} & \dots & 0_{n \times n} \\ 0_{n \times n} & \hat{V}_{0,n \times n} & \dots & \dots \\ \dots & 0_{n \times n} & \dots & \dots \\ \dots & \dots & \dots & \hat{V}_{0,n \times n} \end{bmatrix}, \text{ lumped together to form on big variance-cov matrix}$$

Good start, but...

- The LS solution is still (theoretically) unbiased even if there is residual correlation among repeated measurements on a subject (and design is balanced).
- However, the inference is not unbiased, so must account for correlation when calculating $SE(\hat{\beta})$
- One solution for regularly measured data (e.g., all subjects have same measurements at same times) is to use residuals and estimate the variance-covariance matrix of the data assuming each subject has the same matrix.
- However, can even get more flexible (robust SE' s).

Robust (no Model) SE' s

- Instead of assuming that every subject has the same V_i , one can estimate each subject's V-C matrix separately.
- Thus, the estimate of the variance of the first measurement on subject i is:

$$\hat{v}_{i11} = \hat{\text{var}}(Y_{i1}) = r_{i1}^2 = (Y_{i1} - \hat{Y}_{i1})^2$$

- The correlation between the 1st and 2nd measurement on the i th person is estimated as:
 $\hat{v}_{i12} = \hat{\text{cov}}(Y_{i1}, Y_{i2}) = r_{i1}r_{i2} = (Y_{i1} - \hat{Y}_{i1})(Y_{i2} - \hat{Y}_{i2})$

Robust SE's, cont.

- This procedure is repeated until all the variances and covariances are estimated for each subject (again, we always assume no between subject correlation).
- These are terrible estimators of the variances and covariances.
- However, do we care? No we don't!!!
- Because, they still result, under certain assumptions, in a consistent estimator of $SE(\beta)$
- The estimate, $\text{var}(\hat{\beta})$, sums over the, \hat{V} which means we gain precision by averaging over the individual estimates of the \hat{v}_{ijk}

Get robust inference via Influence Curve (IC)

- Estimators of coefficients in these regression models are so-called asymptotically linear,

$$(\beta - \hat{\beta}) \approx \frac{1}{m} \sum_{i=1}^m IC(Y_i, X_i; \beta) \text{ so}$$

$$\text{var}(\hat{\beta}) = \frac{\text{var}(IC)}{m}$$

- In case of OLS, simple model $E(Y_{ij}|X_{ij}) = \alpha + \beta X_{ij}$

$$IC(Y_i, X_i; \beta) = \frac{1}{E \sum_j X_{ij}^2} \sum_j X_{ij} \{Y_{ij} - (\alpha + \beta X_{ij})\}$$

- Also, *clustered* bootstrap

Estimate linear model for dental data

- ▶ Again, the model we wanted to estimate is:

$$E(Y_{ij} \mid \vec{X}_{ij}) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij1} X_{ij2}$$

Use simple OLS

```
# First, make interaction term
```

```
library(RCurl)
```

```
dat = dat %>% mutate(inter=gender*age)
```

```
ols_fit = lm(distance ~ gender+age+inter, data=dat)
```

```
summary(ols_fit)
```

```
##
```

```
## Call:
```

```
## lm(formula = distance ~ gender + age + inter, data = dat)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -5.6156 -1.3219 -0.1682  1.3299  5.2469
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  17.3727     1.7080  10.171 < 2e-16 ***  
## gender       -1.0321     2.2188  -0.465  0.64279  
## age          0.4795     0.1522   3.152  0.00212 **  
## inter        0.3048     0.1977   1.542  0.12608
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 2.257 on 104 degrees of freedom
```

```
## Multiple R-squared:  0.4227, Adjusted R-squared:  0.4061
```

```
## F-statistic: 25.39 on 3 and 104 DF, p-value: 2.108e-12
```

Output to reporting

- ▶ Because of the interaction term, one can not interpret the individual coefficients as easily as a model with only main terms.
- ▶ The first thing is to decide what is a good measure of association.
- ▶ First lets take age:
 - ▶ must choose a change of age: lets take 6 years
 - ▶ For now, we will Assume that gender is set to some fixed value, g .

$$E(Y_{ij} \mid X_{ij1} = age + 6, X_{ij2} = g) - E(Y_{ij} \mid X_{ij1} = age, X_{ij2} = g)$$

$$= \beta_0 + \beta_1(age + 6) + \beta_2g + \beta_3(age + 6) * g$$

$$- [\beta_0 + \beta_1age + \beta_2g + \beta_3age * g]$$

Getting estimates of linear combination of coefficients

- ▶ The previous slide shows that, for females:

$$\begin{aligned} E(Y_{ij} \mid X_{ij1} = age + 6, X_{ij2} = 1) - E(Y_{ij} \mid X_{ij1} = age, X_{ij2} = 1) \\ = 6 * \beta_1 + 6 * \beta_3 \end{aligned}$$

- Thus, we want to get an estimate of (and standard error for):
 $6 * \hat{\beta}_1 + 6 * \hat{\beta}_3$.

- ▶ We've already done this based on user-written function, but this time we'll use an existing function in the **multcomp** package.

Get estimates and inference for linear combination of coefficients using the “naive” inference returned by the OLS procedure.

```
## Note how the algebra is entered into function
### For boys
lin.boys=glht(ols_fit, linfct = c("6*age +6*inter=0"))
#summary(lin.boys)
confint(lin.boys)
```

```
##
## Simultaneous Confidence Intervals
##
## Fit: lm(formula = distance ~ gender + age + inter, data = dat)
##
## Quantile = 1.983
## 95% family-wise confidence level
##
##
## Linear Hypotheses:
##              Estimate lwr      upr
## 6 * age + 6 * inter == 0 4.7063  3.2051 6.2074
```

Repeat for girls

In this case,

$$E(Y_{ij} \mid X_{ij1} = \text{age} + 6, X_{ij2} = 0) - E(Y_{ij} \mid X_{ij1} = \text{age}, X_{ij2} = 0) \\ = 6 * \beta_1$$

```
lin.girls=glht(ols_fit, linfct = c("6*age=0"))  
#summary(lin.girls)  
confint(lin.girls)
```

```
##  
## Simultaneous Confidence Intervals  
##  
## Fit: lm(formula = distance ~ gender + age + inter, data = dat)  
##  
## Quantile = 1.983  
## 95% family-wise confidence level  
##  
##  
## Linear Hypotheses:  
##           Estimate lwr      upr  
## 6 * age == 0 2.8773  1.0668 4.6877
```

The results suggest that the mean change in the distance for boys is 4.76 (95% 3.20-6.21) and that for girls is 2.88 (95% 1.07-4.69).

Robust estimation of SE's

- ▶ Here, we are beginning the meat of the course, getting inference that can account for the potential repeated measures structure.
- ▶ We will explore many methods that can do this with varying numbers of assumptions
- ▶ Here we start with some of the methods are meant to give **nonparametric inference** for any general linear model estimators.

Inference for sandwich estimator

```
# Read in user written function using lmtest and  
# sandwich packages that return both information  
# on coefficients and estimates of linear  
# combination of coefficients using  
# clustered/robust variance estimates  
source("clx.R")  
source("clx_lincom.R")
```

```
clx(ols_fit, 1, dat$child)
```

```
##  
## t test of coefficients:  
##  
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 17.372727  0.749604 23.1759 < 2.2e-16 ***  
## gender      -1.032102  1.424137 -0.7247  0.47025  
## age         0.479545  0.065257  7.3486 4.712e-11 ***  
## inter       0.304830  0.120799  2.5234  0.01313 *  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Robust estimate for change in mean  
# for 6 year increase in age for  
# boys  
clx.lincom(ols_fit, 1, dat$child, c(0,0,6,6))
```

```
##           Est           CI           pvalue  
## [1,] "4.7063" "3.5108 - 5.9017" "0.0000"
```

Comparison of results

- ▶ If one looks at the age coefficient, one can see that the robust SE is much smaller than the naive one returned by standard OLS: 0.065 vs 0.152.
- ▶ We can also compare the results of the estimate of

$$= 6 * \beta_1 + 6 * \beta_3$$

and we see the CI based upon the robust estimates that account for dependence is 3.511-5.911 versus 3.205-6.207, so for this parameter, the CI gets bigger when the SE is estimated correctly.

Conclusions

- ▶ One can motivate standard ordinary least squares estimation by either maximum likelihood estimation or estimating equation.
- ▶ Ignoring the repeated measures structures does not affect the biasedness of the OLS estimates of the coefficients.
- ▶ However, the standard errors returned assuming independence are biased.
- ▶ We can fix these by procedures that assume nothing about the covariance of observations within the same unit.