Lecture 12

Finish Implementation of WLS/robust variance in R

Introduction to Linear Mixed Models

Weighted least squares review

Assume a longitudinal design with (Y_{ij}, X_{ij}) for i = 1, …, m and j = 1, …, n_i

$$Y_i = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in_i} \end{bmatrix}$$

The model for subject i can be expressed as $Y_i = X_i\beta + \varepsilon_i, \varepsilon_i \sim MVN(0, V_i)$

↳ design matrix for subject i

Then, we can stack the subject models together as $Y = X\beta + \varepsilon, \varepsilon \sim MVN(0, \Sigma)$

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{in_i} \end{bmatrix}$$

where $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_m \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix}, \Sigma = \begin{bmatrix} \boxed{V_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boxed{V_m} \end{bmatrix}$

We have shown how to extend OLS to WLS to account for Σ (instead of OLS assumption of $\boxed{\sigma^2 I}$.

The score equations for the WLS solution is: $\boxed{X'\Sigma^{-1}(Y - X\beta) = 0}$

Yielding: $\hat{\beta}_{wls} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y$ *and* $Var(\hat{\beta}_{wls}) = (X'\Sigma^{-1}X)^{-1}$

Implementation in R

- In Lecture 11, we walked through the required exploratory analysis for longitudinal data. In order to fit a WLS model, we have to understand
  - the mean model ✓
  - the correlation structure ✓
  - the patterns of variance ✓

- Using the simulated NEPAL1 dataset, we settled on:
  - $Y_{ij} = \beta_0 + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + \varepsilon_{ij}, \varepsilon_{ij} \sim Normal$
  - $Corr(\varepsilon_{ij}, \varepsilon_{ik}) = \rho^{|j-k|}$ AR1 model
  - $Var(\varepsilon_{ij}) = f(age_{ij})$

- Fit this model using gls, but also considered two other correlation models:
  - $Corr(\varepsilon_{ij}, \varepsilon_{ik}) = \rho$ exchangeable / compound symmetry
  - $Corr(\varepsilon_{ij}, \varepsilon_{ik}) = \rho_{jk}$ → unstructured

Implementation in R

- AR1 model, constant variance
  - mod.gls.exch.het = gls(wt ~ age + age_sp6, data = nepal1,correlation = corAR1(form = ~num|id))
- Exchangeable / compound symmetry, constant variance
  - mod.gls.exch.het = gls(wt ~ age + age_sp6, data = nepal1, correlation = corCompSymm(form = ~1|id))
- Unstructured, constant variance
  - mod.gls.exch.het = gls(wt ~ age + age_sp6, data = nepal1, correlation = corSymm(form = ~num|id))
- Allowing for variance to depend on age
  - mod.gls.exch.het = gls(wt ~ age + age_sp6, data = nepal1,correlation = corAR1(form = ~num|id), weights = varFunc(~age))

$\hookrightarrow \log(\sigma^2) = \gamma_0 + \gamma_1 age$

Gamma model

* need to pass information about the order of the data
→ specify a variable
→ sort the data

Id, num

Generalized Estimating Equations

Weighted least squares is a special case of a general method called Generalized Estimating Equations (GEE).

In the case of $Y_i \sim MVN(X_i\beta, V_i)$, the WLS/GEE method finds the values of β that equates the score equations (i.e. estimating equations) to 0. In the case of independent Y_i, for $i = 1, ..., m$, the $\hat{\beta}_{wls}$ solves:

$$S(\beta,\theta) = \sum_{i=1}^{m} \overset{X'}{\frac{\partial X_i\beta}{\partial\beta}} \overset{\Sigma^{-1}}{V_i(\theta)^{-1}} \overset{Y-X\beta}{(Y_i - X_i\beta)} = 0$$

$X'\Sigma^{-1}(Y-X\beta)=0$

↳ WLS

The estimation procedure is iterative, same as WLS.

- We will discuss GEE again in 654.
- One advantage of GEE is that you don't have to specify a multivariate distribution for Y_i, so long as we can specify $E(Y_{ij}), Var(Y_{ij}), Corr(Y_{ij}, Y_{ik})$, then we can solve for β and make inferences.
  - This is nice because multivariate Bernoulli or Poisson distributions are quite complicated.

Generalized Estimating Equations

- Why do we care about GEE?

- Historical: Kung-Yee Liang and Scott Zeger derived the method; motivated by a longitudinal design with binary outcome

- The *gls* function in R is limiting in that it does not directly compute robust variance estimates; so you only have access to standard error estimates based on the model you specify.
  - See *clubSandwich* which should work on *gls* objects and produce robust variance estimates; I have had trouble with this function

- The typical implementation of GEE is to provide both model based (similar to *gls*) and robust variance estimates.

- Note: in *gee* function in R, you specify a model for the within subject/cluster correlation structure (R_i) and the model assumes a constant variance

Robust Variance Estimation

- We refit several of the models we considered before (plus a few additional models) and obtained robust variance estimates for some of the models
  - OLS with and without robust variance estimate
  - Exchangeable, constant variance with and without robust variance estimate
  - Exchangeable, variance depends on age
  - AR1, constant variance with and without robust variance estimate
  - AR1, variance depends on age

GEE Naive robust

Lab 6

$\hat{\beta}$ $se(\hat{\beta})$

	OLS	OLS-RV	Exch	Exch-Het	Exch-RV
Intercept	5.074 (0.289)	5.074 (0.157)	4.915 (0.199)	5.116 (0.082)	4.916 (0.192)
Age	0.486 (0.06)	0.486 (0.026)	0.512 (0.029)	0.468 (0.025)	0.511 (0.032)
Age_SP1	-0.344 (0.071)	-0.344 (0.028)	-0.366 (0.034)	-0.314 (0.03)	-0.366 (0.034)

	OLS	OLS-RV	AR1	AR1-Het	AR1-RV
Intercept	5.074 (0.289)	5.074 (0.157)	4.98 (0.19)	5.106 (0.073)	4.99 (0.165)
Age	0.486 (0.06)	0.486 (0.026)	0.495 (0.022)	0.467 (0.024)	0.494 (0.023)
Age_SP1	-0.344 (0.071)	-0.344 (0.028)	-0.348 (0.025)	-0.313 (0.028)	-0.347 (0.023)

Which model is “best”?

- You can use an information criteria statistic, which combines information about the fit (i.e. sums of squares residuals) and the complexity of the model (i.e. number of parameters in the model, including the parameters for variance/covariance)

- Akaike’s Information Criteria: -2 log-likelihood + 2 x p, where p is the number of parameters in the model

- Models with smaller AIC values are “better”

```
##                      df      AIC
## mod.gls.exch.fit      5 729.3473
## mod.gls.exch.het.fit  5 827.3266
## mod.gls.ar1.fit       5 589.6508
## mod.gls.ar1.het.fit   5 731.3010
```

Two approaches for modeling longitudinal data

- Descriptive: Marginal model, goal is to describe and make inference for the mean model. ↳ estimate β from $E(Y) = X\beta$
  - Have to account for the variance/correlation structure to get valid inferences
  - But we don't necessary care about describing that structure.

- Etiologic: Conditional models: we are specifically interested in describing where the correlation comes from.
  - E.g. the current observation may depend on the prior observation (transition model)
  - E.g. each subject may be distinguished by latent variables/random effects which separate their data from other subjects data.
  - The goal is to describe the population level patterns (similar to marginal models) but also quantify heterogeneity across subjects in features of the data that are very important for public health researchers, e.g. variation in child specific growth rates.

Transition Models

Here past observations of the outcome cause future values of the outcomes. Namely, a transition model where the current value of Y_{ij} depends on the p past observations can be expressed as:

$$E(Y_{ij}|Y_{ij-1}, ..., Y_{ij-p}, X_{ij}) = X'_{ij}\beta^c + \sum_{k=1}^{p} \alpha_k Y_{ij-k}$$

α_k measures of correlation

The special case of the AR-1 model is where $\boxed{p = 1.}$

$$E(Y_{ij}|Y_{ij-1}, X_{ij}) = X'_{ij}\beta^c + \alpha Y_{ij-1}$$

Note that the models above make a strong assumption: the relationship between the mean of Y_{ij} and X_{ij} is the same regardless of the past values of Y. This assumption can be made flexible by including interaction terms of components of X_{ij} and past values of Y.

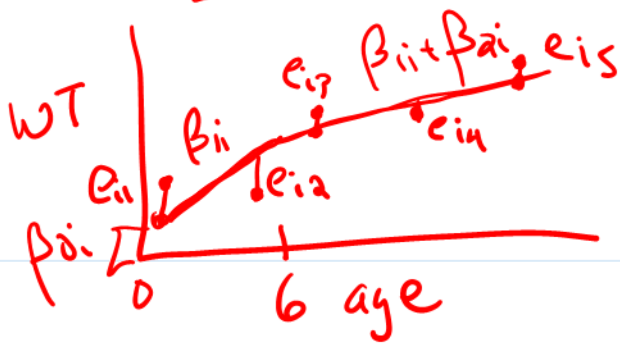
Subject specific or random effects models or mixed models
linear mixed model

- Consider the data generating structure within the NEPAL1 and NEPAL2 simulated datasets:
  - Children are enrolled between 1 and 5 months of age
  - Children are followed over time and growth in weight is recorded every 4 months for a total of 5 assessments (enrollment + 4 follow-ups)

- For each child, we can think of the child's growth:

Child specific model

$$Y_{ij} = \beta_{0i} + \beta_{1i}age_{ij} + \beta_{2i}(age_{ij} - 6)^{+} + e_{ij}$$



Subject specific or random effects models

- The β describe characteristics of the specific children and we assume that these characteristics can vary from child to child, specifically

expected birth weight for child i

pop avg birthweight

residuals

random effects

$$\underbrace{\begin{bmatrix}\boxed{\beta_{0i}}\\ \beta_{1i}\\ \beta_{2i}\end{bmatrix}}_{\beta_i} = \underbrace{\begin{bmatrix}\underline{\beta_0}\\ \beta_1\\ \beta_2\end{bmatrix}}_{\text{pop mean parameters}} + \begin{bmatrix} b_{0i}\\ b_{1i}\\ b_{2i}\end{bmatrix}$$

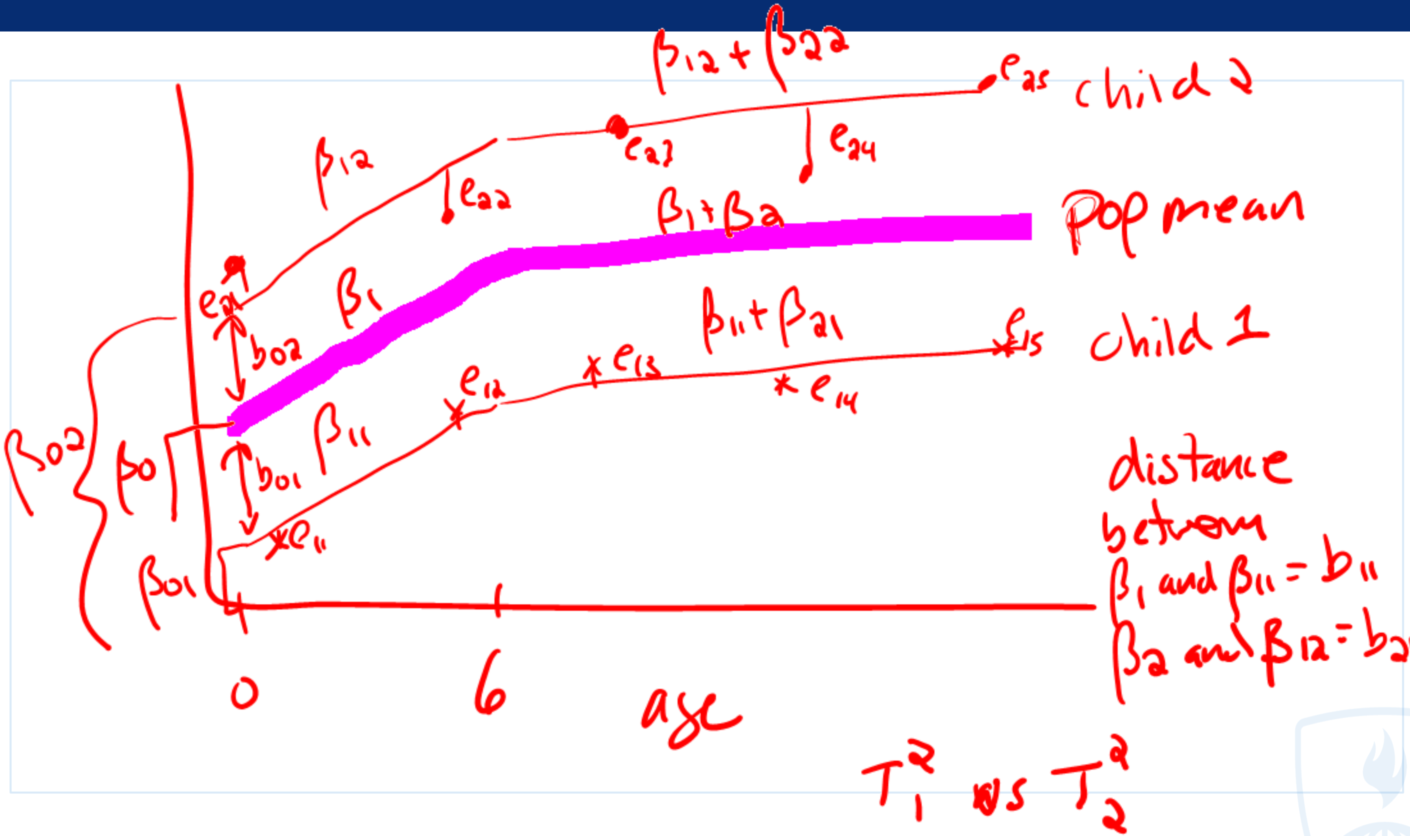
$$\underline{\beta_i} = \underbrace{\beta}_{\text{pop mean fixed effects}} + \underbrace{b_i}_{\text{random effects}}, b_i \sim \underline{MVN(0, D)}, D = \begin{bmatrix}\tau_0^2 & \tau_{01} & \tau_{02}\\ \tau_{01} & \tau_1^2 & \tau_{12}\\ \tau_{02} & \tau_{12} & \tau_2^2\end{bmatrix}$$

$Var(b_{0i}) = \underline{\tau_0^2},$

$Var(b_{1i}) = \tau_1^2$ etc

$Cov(b_{0i}, b_{1i}) = \tau_{01}$

Visualization



General Model

We can rewrite the model above as:

$$Y_{ij} = (\overset{\beta_{0i}}{\beta_0 + b_{0i}}) + (\overset{\beta_{1i}}{\beta_1 + b_{1i}})age_{ij} + (\overset{\beta_{2i}}{\beta_2 + b_{2i}})(age_{ij} - 6)^+ + e_{ij}$$

In vector notation,

$$Y_{ij} = \begin{bmatrix} 1 \\ age_{ij} \\ (age_{ij} - 6)^+ \end{bmatrix}' \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}_{\text{fixed effects}} + \begin{bmatrix} 1 \\ age_{ij} \\ (age_{ij} - 6)^+ \end{bmatrix}' \underbrace{\begin{bmatrix} b_{0i} \\ b_{1i} \\ b_{2i} \end{bmatrix}}_{\text{random effects}} + e_{ij}$$

Even more generally,

$$Y_{ij} = X_{ij}'\beta + \underbrace{Z_{ij}'b_i}_{\text{variation between subjects}} + \underbrace{e_{ij}}_{\text{variation within subject}}$$

where $b_i \sim MVN(0, D)$, e_{ij} iid $N(0, \sigma^2)$ **and** b_i and e_{ij} are independent!

Means and Variances

- In the random effects model, we express the mean function for an individual subject as:

$$E(Y_{ij}|X_{ij}, b_i) = X_{ij}\beta + Z_{ij}b_i$$

- We can express the population mean (i.e. the average over all subjects) as:

law of total expectation

$$E(Y_{ij}|X_{ij}) = \underset{b_i}{E}[E(Y_{ij}|X_{ij}, b_i)] = E[X_{ij}\beta + Z_{ij}b_i] = X_{ij}\beta$$

- We can derive the variance of Y_{ij} as

law of total variance

$$Var(Y_{ij}|X_{ij}) = E_{b_i}[\overbrace{Var(Y_{ij}|X_{ij}, b_i)}^{Var(e_{ij})}] + Var_{b_i}[E(Y_{ij}|X_{ij}, b_i)]$$

$$Var(Y_{ij}|X_{ij}) = E_{b_i}[\sigma^2] + Var_{b_i}[X'_{ij}\beta + Z'_{ij}b_i]$$

$$Var(Y_{ij}|X_{ij}) = \sigma^2 + Z'_{ij}DZ_{ij}$$

within subject heterogeneity

between subject heterogeneity

Correlation

- Assume a random intercept only model: $\Rightarrow$ exchangeable correlation structure
  - $Y_{ij} = \beta_{0i} + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, \beta_{0i} \sim N(\beta_0, \tau_0^2), e_{ij} \overset{iid}{\sim} N(0, \sigma^2), Cov(\beta_{0i}, e_{ij}) = 0$
  - $Y_{ij} = \beta_0 + b_{0i} + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, b_{0i} \sim N(0, \tau_0^2), e_{ij} \sim N(0, \sigma^2), Cov(b_{0i}, e_{ij}) = 0$

- What is $Cov(Y_{ij}, Y_{ik})$? $Cov(\beta_0 + b_{0i} + \beta_1 age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, \beta_0 + b_{0i} + \beta_1 age_{ik} + \beta_2 (age_{ik} - 6)^+ + e_{ik})$

$$= Cov(b_{0i} + e_{ij}, b_{0i} + e_{ik})$$

$$= Cov(b_{0i}, b_{0i}) + \underbrace{\cancel{Cov(b_{0i}, e_{ik})}}_{0} + \underbrace{\cancel{Cov(e_{ij}, b_{0i})}}_{0} + \underbrace{\cancel{Cov(e_{ij}, e_{ik})}}_{0}$$

$$= Var(b_{0i}) = \tau_0^2$$

$$Corr(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\tau_0^2 + \sigma^2} = \frac{\tau_0^2}{\tau_0^2 + \sigma^2} = \rho$$

$$Var(b_{0i} + e_{ij}) = \tau_0^2 + \sigma^2$$

Correlation

- Assume a random intercept and random slope for age model:

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{i1})age_{ij} + \beta_2 (age_{ij} - 6)^+ + e_{ij}, \textit{ where}$$

$$b_{0i} \sim N(0, \tau_0^2), b_{1i} \sim N(0, \tau_1^2), Cov(b_{0i}, b_{1i}) = \tau_{01}, e_{ij} \sim N(0, \sigma^2), Cov(b_{0i}, e_{ij}) = 0, Cov(b_{1i}, e_{ij}) = 0$$

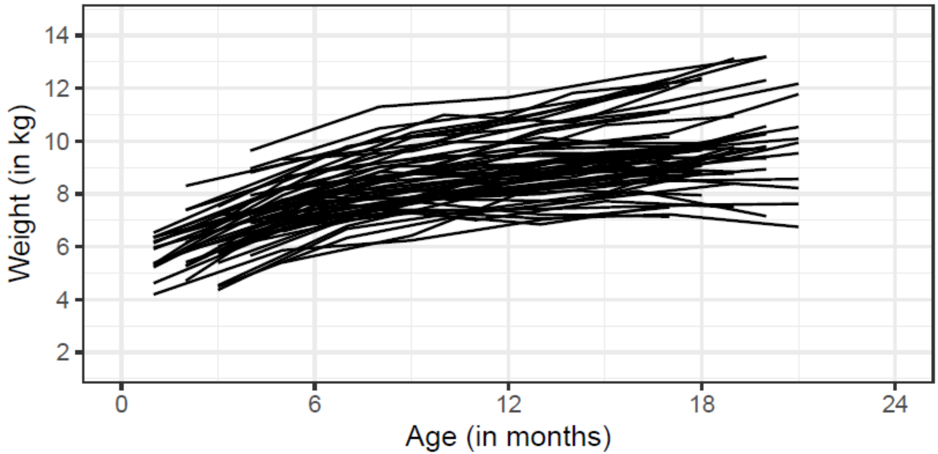
$$Cov(Y_{ij}, Y_{ik}) = Cov(b_{0i} + b_{1i} age_{ij} + e_{ij},\ b_{0i} + b_{1i} age_{ik} + e_{ik})$$

$$= Cov(b_{0i}, b_{0i}) + Cov(b_{0i}, b_{1i} age_{ik}) + Cov(b_{1i} age_{ij}, b_{0i}) + Cov(b_{1i} age_{ij}, b_{1i} age_{ik})$$

$$= Var(b_{0i}) + age_{ik} Cov(b_{0i}, b_{1i}) + age_{ij} Cov(b_{1i}, b_{0i}) + age_{ij} age_{ik} Var(b_{1i})$$

$$= \tau_0^2 + \underbrace{age_{ik}}\tau_{01} + \underbrace{age_{ij}}\tau_{01} + \underbrace{age_{ij} age_{ik}}\tau_1^2$$

Example: NEPAL1 simulated data



spaghetti

- Fit the following model to the NEPAL1 simulated dataset

$$Y_{ij} = \beta_0 + b_{0i} + (\beta_1 + b_{i1})age_{ij} + \beta_2(age_{ij} - 6)^+ + e_{ij},\ where$$

$$b_{0i} \sim N(0, \tau_0^2), b_{1i} \sim N(0, \tau_1^2), Cov(b_{0i}, b_{1i}) = \tau_{01}, e_{ij} \sim N(0, \sigma^2), Cov(b_{0i}, e_{ij}) = 0, Cov(b_{1i}, e_{ij}) = 0$$

Z → how is data clustered

```
lmer(wt~age+age_sp6+(1+age|id),data=nepal2, control = lmerControl(optimizer ="Nelder_Mead"))
```

y X random effects

Example: NepalI simulated data

- What is the estimate of the population mean birth weight?

For the pop of Nepali children the average birthweight is 5 kg

- What is the estimate of the population mean growth rate in the first 6 months of life?

On average, children's weights increase by .5 kg per month during the 1st 6 months of life.

pop mean parameter estimates

```
##              Estimate  Std. Error    t value
## (Intercept) 4.9777731  0.15426618   32.26743
## age         0.4984283  0.01867078   26.69563
## age_sp6    -0.3497761  0.01802296  -19.40725
```

$\hat\beta_0$ $\hat\beta_1$ $\hat\beta_2$

```
summary(fit)$varcor
```

```
##  Groups   Name        Std.Dev. Corr
##  id       (Intercept) 1.045047
##           age         0.082252 -0.345
##  Residual             0.281274
```

b_{0i} b_{1i} $\hat\tau_0$ $\hat\tau_1$ $\hat\sigma$

$$\frac{\hat\tau_{01}}{\hat\tau_0 \cdot \hat\tau_1}$$

```
est = fixef(fit)
```

- What is the estimate of the difference in the population mean growth rate after 6 months compared to during the first 6 months of life?

-.35 kg per month

Example: NepalI simulated data

- For a given child at a specific age, how much do the observed weights differ (+/-) on average from the child's average weight at that age?

For a given child, the average distance between their observed wt and their predicted wt is .28 kg

```
##               Estimate Std. Error   t value
## (Intercept)  4.9777731 0.15426618  32.26743
## age          0.4984283 0.01867078  26.69563
## age_sp6     -0.3497761 0.01802296 -19.40725
```

```r
summary(fit)$varcor
```

```
##  Groups   Name        Std.Dev. Corr
##  id       (Intercept) 1.045047
##                       age         0.082252 -0.345
##  Residual             0.281274
```

$\hat{\sigma}$ ← within child residual

```r
est = fixef(fit)
```

- Construct an interval that contains 95% of birthweights for Nepali children.

β_{0i} = child specific birthweights

$\hat{\tau}_0 = 1.05$

β_0 pop avery birthweight

$\hat{\beta}_0 +/- 1.96\, \hat{\tau}_0$

$5 +/- 2 \times 1.05$

rounding: 3 to 7 kg

Example: NepalI simulated data

- Construct an interval that contains 95% of growth rates for Nepali children under 6 months of age

pop average growth for 1st 6 months .5 kgs per month

β_{1i}: child specific growth rate in 1st 6 months

$\hat{\tau}_1 = .08$

```
##              Estimate Std. Error   t value
## (Intercept) 4.9777731 0.15426618  32.26743
## age         0.4984283 0.01867078  26.69563
## age_sp6    -0.3497761 0.01802296 -19.40725
```

```
summary(fit)$varcor
```

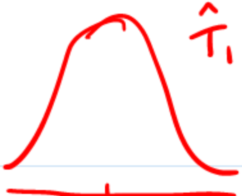
```
##  Groups   Name        Std.Dev. Corr
##  id       (Intercept) 1.045047
##           age         0.082252 -0.345
##  Residual             0.281274
```

```
est = fixef(fit)
```

$.5 \pm 2(.08)$

.34 to .66 kgs per month

- Construct an interval that contains 95% of growth rates for Nepali children over 6 months of age



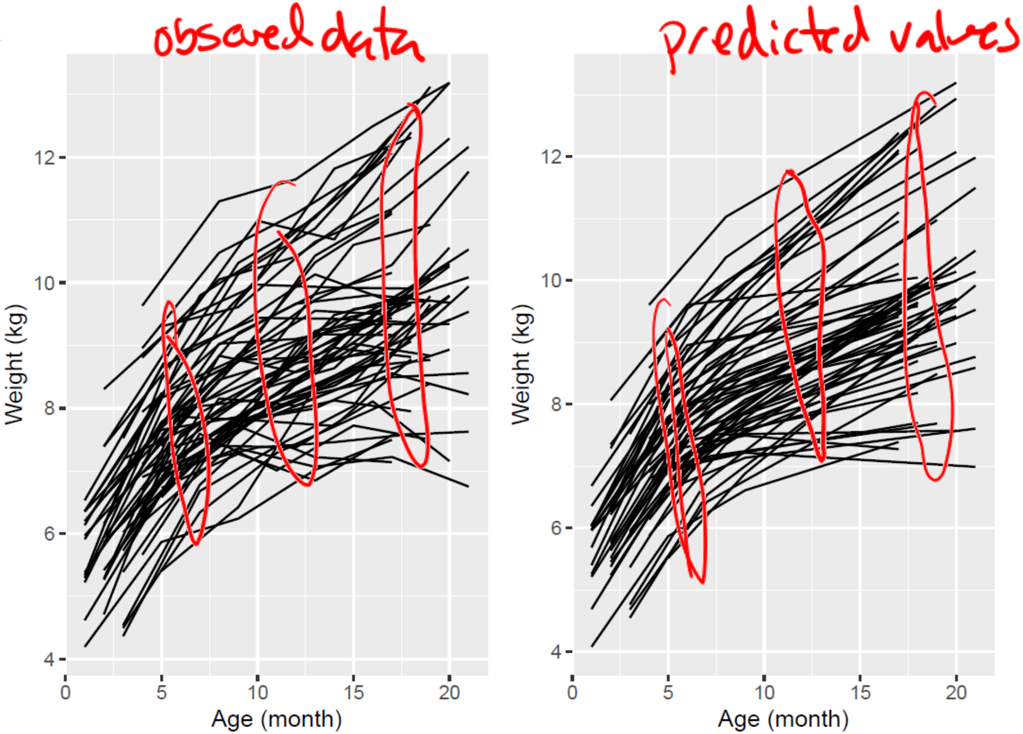
$\hat{\tau}_1 = .08$

$\hat{\beta}_1 + \hat{\beta}_2 = .5 - .35 = .15$

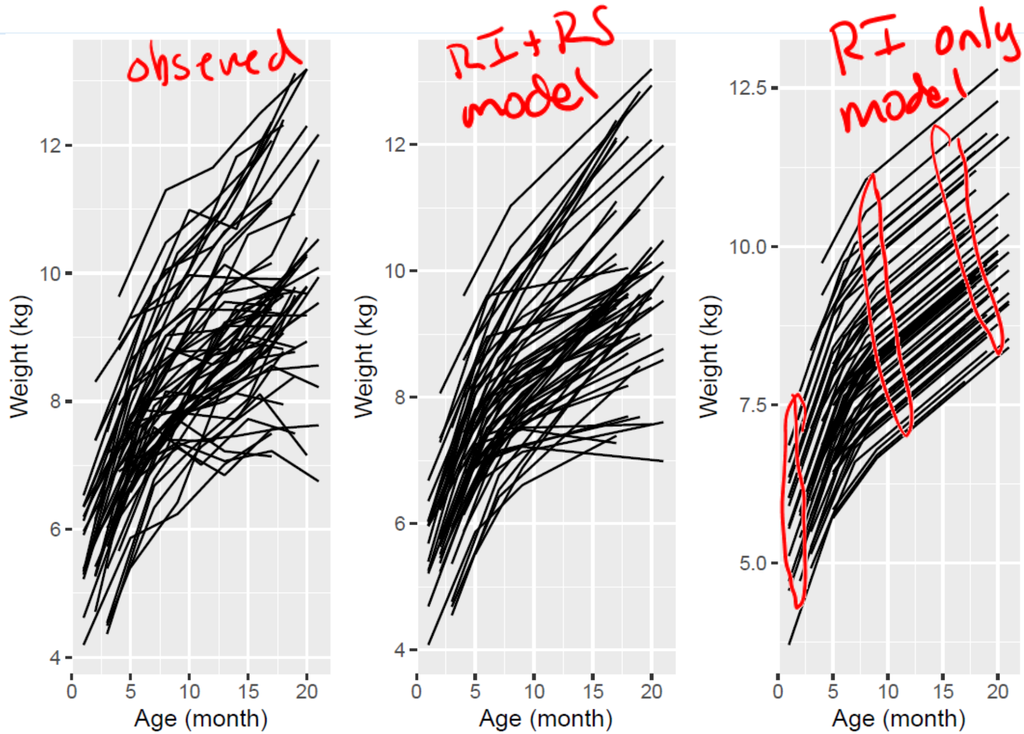
$.15 \pm 2(.08)$

-.01 to .31 kgs per month

Example: NepaliI simulated data



Example: NepaliI simulated data



Information criterion comparison

10. Compare the fits from the *gls* models and random intercept and slope models, using AIC.

```
AIC(mod.gls.exch.fit,mod.gls.exch.het.fit,mod.gls.ar1.fit,mod.gls.ar1.het.fit)
```

```
##                      df      AIC
## mod.gls.exch.fit      5 729.3473
## mod.gls.exch.het.fit  5 827.3266
## mod.gls.ar1.fit       5 589.6508
## mod.gls.ar1.het.fit   5 731.3010
```

```
AIC(fit,fit.int)
```

```
##         df      AIC
## fit      7 526.8088
## fit.int  5 729.3473
```

NOTE: The data was generated under the random intercept and random slope for age model!

Next time....

- On Tuesday March 9th, we will review
  - linear mixed models
  - analyze NEPAL2 together

Lecture 12 Handout
Rmd

▷ Lab 7 and 8 are open sessions