

Advanced Methods in Biostatistics II

Lecture 12

December 5, 2017

General linear mixed model

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

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \varepsilon_i, \quad i = 1, \dots, m.$$

- Here:
 - \mathbf{y}_i is an $n_i \times 1$ vector of observations.
 - \mathbf{X}_i is an $n_i \times p$ design matrix for the fixed effects.
 - $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects.
 - \mathbf{Z}_i is an $n_i \times k$ design matrix for the random effects.
 - \mathbf{u}_i is a $k \times 1$ vector of random effects.
 - ε_i is an $n_i \times 1$ vector of error terms.

General linear mixed model

- In addition, we assume that the error terms and random effects are uncorrelated across groups, i.e., when $i \neq i'$:

$$\text{cov}(\varepsilon_i, \varepsilon_{i'}) = \mathbf{0}, \quad \text{cov}(\mathbf{u}_i, \mathbf{u}_{i'}) = \mathbf{0}, \quad \text{cov}(\varepsilon_i, \mathbf{u}_{i'}) = \mathbf{0}.$$

- When $i = i'$, we have that

$$\text{cov}(\varepsilon_i, \varepsilon_i) = \Sigma_\varepsilon, \quad \text{cov}(\mathbf{u}_i, \mathbf{u}_i) = \Sigma_u, \quad \text{cov}(\varepsilon_i, \mathbf{u}_i) = \mathbf{0}.$$

General linear mixed model

- We can reformulate the model as follows:

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

where

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_m \end{pmatrix} \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{Z}_m \end{pmatrix}$$

$$\mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_m \end{pmatrix} \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix}.$$

General linear mixed model

- The model implies that

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

where

$$\mathbf{V} = \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}' + \boldsymbol{\Sigma}_\epsilon.$$

- This is called the marginal formulation of the model.
- Estimation and inference for the fixed effects parameters are based on this formulation.

General linear mixed model

- \mathbf{V} is an $N \times N$ block-diagonal matrix, with $N = \sum_{i=1}^m n_i$.
- It can be written as follows:

$$\mathbf{V} = \begin{pmatrix} \mathbf{Z}_1 \Sigma_u \mathbf{Z}_1' + \Sigma_\epsilon & \mathbf{0} & \vdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \Sigma_u \mathbf{Z}_2' + \Sigma_\epsilon & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{Z}_m \Sigma_u \mathbf{Z}_m' + \Sigma_\epsilon \end{pmatrix}$$

- A key feature of mixed models is that it is possible to predict the unobserved vector \mathbf{u} .
- The Best Linear Unbiased Predictor (BLUP) for \mathbf{u} is given by $\hat{\mathbf{u}} = E(\mathbf{u}|\mathbf{y})$.
- When \mathbf{u} and \mathbf{y} are joint multivariate normal this can be expressed as follows:

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

Example

- Let us revisit the pig example.
- It consists of 9 repeated weight measures on 48 pigs.
- We use the following model:

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

for $i = 1, \dots, 48$ and $j = 1, \dots, 9$ where $u_i \sim_{iid} N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma_\epsilon^2)$.

Example

- For the pig data set:

$$\mathbf{y} = (y_{1,1} \ \dots \ y_{1,9} \ \dots \ y_{48,1} \ \dots \ y_{48,9})'$$

$$\mathbf{x} = \begin{pmatrix} 1 & \dots & 1 & \dots & 1 & \dots & 1 \\ 1 & \dots & 9 & \dots & 1 & \dots & 9 \end{pmatrix}'$$

$$\boldsymbol{\beta} = (\beta_0 \ \beta_1)'$$

$$\mathbf{Z} = \mathbf{I}_{48} \otimes \mathbf{J}_9 \quad \mathbf{u} = (u_1 \ \dots \ u_{48})'$$

$$\Sigma_u = \sigma_u^2 \mathbf{I}_9 \quad \Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}_N$$

Example

- We can predict u_i by considering the estimate $E[u_i | \mathbf{y}]$.
- To derive this, note the density for $u_i | \mathbf{y}$ is equal to the density of $u_i | \mathbf{y}_i$, 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

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

and

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

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.

- Consider expressing the model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{(1)}\mathbf{u}_1 + \cdots + \mathbf{Z}_{(k)}\mathbf{u}_k + \boldsymbol{\varepsilon}$$

- Here we assume each $\mathbf{Z}_{(h)}$ is an $n \times r_h$ matrix that specifies membership in the various clusters or subgroups, and the \mathbf{u}_h are different random effects.

- Let $E(\mathbf{u}) = E(\varepsilon) = \mathbf{0}$, and let us assume

$$\text{var}(\varepsilon) = \sigma_\varepsilon^2 \mathbf{I}_N$$

and

$$\text{var}(\mathbf{u}_h) = \sigma_h^2 \mathbf{I}_{r_h}.$$

- Here r_h represents the number of elements in \mathbf{u}_h .
- In addition, $\text{cov}(\mathbf{u}_i, \mathbf{u}_j) = \mathbf{0} \ \forall i \neq j$, and $\text{cov}(\mathbf{u}, \varepsilon) = \mathbf{0}$.

Linear mixed models

- We summarize the model as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{h=1}^k \mathbf{Z}_{(h)}\mathbf{u}_h + \boldsymbol{\varepsilon}$$

where

$$\mathbf{V} \equiv \text{var}(\mathbf{y}) = \sum_{h=1}^k \sigma_h^2 \mathbf{Z}_{(h)} \mathbf{Z}_{(h)}' + \sigma_\epsilon^2 \mathbf{I}_N.$$

Linear mixed models

- A useful extension is to make the following definition:

$$\mathbf{u}_0 \equiv \varepsilon \quad \mathbf{Z}_{(0)} \equiv \mathbf{I}_N \quad \sigma_0^2 \equiv \sigma_\varepsilon^2$$

- We can now write the model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{h=0}^k \mathbf{Z}_{(h)} \mathbf{u}_h$$

with

$$\mathbf{V} = \sum_{h=0}^k \sigma_h^2 \mathbf{Z}_{(h)} \mathbf{Z}_{(h)}'$$

- Next, we focus on estimating the fixed effects and variance components.
- For a given \mathbf{V} we can estimate β using:

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

- The most common approaches towards estimating the parameters in the covariance matrices is to use maximum likelihood (ML) or restricted maximum likelihood (REML).

- The MLE of \mathbf{V} is based on using the marginal model:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V}).$$

- The log-likelihood of \mathbf{y} is given by:

$$\ell(\boldsymbol{\beta}, \mathbf{V}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{V}| - \frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

- Optimizing over β for fixed \mathbf{V} we obtain the familiar estimate:

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

- Plugging this value into $\ell(\beta, \mathbf{V})$ we obtain the profile log-likelihood for \mathbf{V} :

$$\begin{aligned}\ell_P(\mathbf{V}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta}) \\ &\propto -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}(\mathbf{y}'\mathbf{V}^{-1}(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1})\mathbf{y})\end{aligned}$$

- The MLE for \mathbf{V} can be found by maximizing this function.
- Note that the MLE estimator for the linear mixed effect model will be biased and the error on the random effects covariance may be large.
- In contrast, REML estimates tend to be less biased than the ML estimates.

- For example, if $y_i \sim_{iid} N(\mu, \sigma^2)$, REML yields the unbiased variance estimate (divided by $n - 1$) rather than the biased variance estimate (divided by n) obtained via ML.
- REML estimates are often the default for linear mixed effect model programs.

- The key idea of REML is to perform maximum likelihood estimation for \mathbf{Ky} rather than \mathbf{y} , where \mathbf{K} is chosen so that the distribution of \mathbf{Ky} only involves the variance components, not β .
- For this to occur, we need \mathbf{K} to be a full-rank matrix such that $\mathbf{KX} = \mathbf{0}$.
- In this setting, $E(\mathbf{Ky}) = \mathbf{KX}\beta = \mathbf{0}$.

- It turns out that \mathbf{K} must be of the form

$$\begin{aligned}\mathbf{K} &= \mathbf{C}(\mathbf{I} - \mathbf{H}) \\ &= \mathbf{C}[\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'],\end{aligned}$$

where \mathbf{C} specifies a full-rank transformation of the rows of the projection matrix $\mathbf{I} - \mathbf{H}$.

- There are an infinite number of such \mathbf{K} 's, and it does not matter which is used.

- Consider the marginal model where

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

with

$$\mathbf{V} = \sum_{h=0}^k \sigma_h^2 \mathbf{z}_{(h)} \mathbf{z}_{(h)}'.$$

- Let \mathbf{K} be specified as above.
- Then,

$$\mathbf{Ky} \sim N\left(\mathbf{0}, \mathbf{K}\left(\sum_{h=0}^k \sigma_h^2 \mathbf{z}_{(h)} \mathbf{z}_{(h)}'\right) \mathbf{K}'\right).$$

- Thus the distribution of **Ky** only involves the $k + 1$ variance components as unknown parameters.
- In order to estimate the variance components, the next step in REML is to maximize the likelihood of **Ky** with respect to these variance components.
- We now develop a set of estimating equations by taking partial derivatives of the log likelihood with respect to the variance components, and setting them equal to zero.

- The log-likelihood can be expressed as follows:

$$\begin{aligned}
 \ell(\sigma_0^2, \dots, \sigma_k^2) &= -\frac{n-r}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{K}\mathbf{V}\mathbf{K}'| - \frac{1}{2} \mathbf{y}'\mathbf{K}'(\mathbf{K}\mathbf{V}\mathbf{K}')^{-1} \mathbf{K}\mathbf{y} \\
 &= -\frac{n-r}{2} \log(2\pi) - \frac{1}{2} \log \left| \mathbf{K} \left(\sum_{h=0}^k \sigma_h^2 \mathbf{z}_{(h)} \mathbf{z}_{(h)}' \right) \mathbf{K}' \right| \\
 &\quad - \frac{1}{2} \mathbf{y}'\mathbf{K}' \left[\mathbf{K} \left(\sum_{h=0}^k \sigma_h^2 \mathbf{z}_{(h)} \mathbf{z}_{(h)}' \right) \mathbf{K}' \right]^{-1} \mathbf{K}\mathbf{y}
 \end{aligned}$$

Vector and Matrix Calculus

Theorem

Let \mathbf{A} be a nonsingular with a derivative $\partial\mathbf{A}/\partial x$. Then,

$$\frac{\partial\mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1} \frac{\partial\mathbf{A}}{\partial x} \mathbf{A}^{-1}$$

Vector and Matrix Calculus

Theorem

Let \mathbf{A} be a positive definite matrix. Then,

$$\frac{\partial \log |\mathbf{A}|}{\partial \mathbf{x}} = \text{tr} \left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{x}} \right).$$

- Using these results, we take the partial derivative of $\ell(\sigma_0^2, \dots, \sigma_k^2)$ with respect to each of the σ_h^2 :

$$\begin{aligned}
 & \frac{\partial}{\partial \sigma_h^2} \ell(\sigma_0^2, \dots, \sigma_k^2) \\
 = & -\frac{1}{2} \text{tr} \left((\mathbf{KVK}')^{-1} \left[\frac{\partial}{\partial \sigma_h^2} (\mathbf{KVK}') \right] \right) \\
 & + \frac{1}{2} \mathbf{y}' \mathbf{K}' (\mathbf{KVK}')^{-1} \left[\frac{\partial}{\partial \sigma_h^2} (\mathbf{KVK}') \right] (\mathbf{KVK}')^{-1} \mathbf{Ky} \\
 = & -\frac{1}{2} \text{tr} \left((\mathbf{KVK}')^{-1} \mathbf{KZ}_{(h)} \mathbf{Z}_{(h)}' \mathbf{K}' \right) \\
 & + \frac{1}{2} \mathbf{y}' \mathbf{K}' (\mathbf{KVK}')^{-1} \mathbf{KZ}_{(h)} \mathbf{Z}_{(h)}' \mathbf{K}' (\mathbf{KVK}')^{-1} \mathbf{Ky}
 \end{aligned}$$

- Setting this result equal to zero, we obtain a set of $k + 1$ estimating equations for $\sigma_0^2, \dots, \sigma_k^2$ given by

$$\begin{aligned} & \text{tr} \left((\mathbf{KVK}')^{-1} \mathbf{KZ}_{(h)} \mathbf{Z}_{(h)}' \mathbf{K}' \right) \\ &= \mathbf{y}' \mathbf{K}' (\mathbf{KVK}')^{-1} \mathbf{KZ}_{(h)} \mathbf{Z}_{(h)}' \mathbf{K}' (\mathbf{KVK}')^{-1} \mathbf{Ky} \end{aligned}$$

- Note, the expected value of the quadratic form on the right side is given by the left side of the equation.

- In certain special cases these equations can be simplified to yield closed-form solutions.
- However, in most cases, numerical methods are required to solve the system of equations.

Bayesian formulation

- An alternative way to derive the REML estimates is to use a Bayesian formulation.
- Consider a model where

$$\mathbf{y} \mid \boldsymbol{\beta} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\boldsymbol{\Sigma}_u\mathbf{Z}' + \sigma^2\mathbf{I})$$

and $\boldsymbol{\beta} \sim N(0, \theta\mathbf{I})$.

- Calculating the posterior mode for $\boldsymbol{\Sigma}_u$ and σ^2 after integrating out $\boldsymbol{\beta}$ as $\theta \rightarrow \infty$ results in the REML estimates.

- Estimates of the variance components can be inserted into \mathbf{V} to obtain:

$$\hat{\mathbf{V}} = \sum_{h=0}^k \hat{\sigma}_h^2 \mathbf{z}_{(h)} \mathbf{z}_{(h)}'.$$

- We can in turn use this to estimate:

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

- This is sometimes called the estimated generalized least-squares solution.

- An approximate estimate of the variance-covariance matrix for $\hat{\beta}$ is given by:

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

- Note this ignores the variability due to the estimation of \mathbf{V} .
- For large samples this extra variability is negligible, but it can be substantial for smaller samples.
- If the sample size is small it may be preferable to use a parametric bootstrap to approximate the distribution of the test statistics.

Likelihood ratio tests

- Testing can be performed using likelihood ratio tests.
- We can compare two nested models m_0 and m_1 using the likelihood ratio test statistic:

$$-2\log(LR(\mathbf{y})) = -2(\ell(\hat{\beta}_0, \hat{\mathbf{V}}_0|\mathbf{y}) - \ell(\hat{\beta}_1, \hat{\mathbf{V}}_1|\mathbf{y})),$$

where $\hat{\beta}_0$, $\hat{\mathbf{V}}_0$ and $\hat{\beta}_1$, $\hat{\mathbf{V}}_1$ are the parameter estimates under the two models.

- This test statistic is asymptotically χ^2 with degrees of freedom equal to the difference in number of parameters between the two models.

Likelihood ratio tests

- If the models differ in their fixed effects, it is not possible to use REML estimates in the likelihood ratio statistic.
- REML estimates the random effects by considering linear combinations of the data that remove the fixed effects and therefore the two likelihood functions will not be comparable.

Likelihood ratio tests

- Note this asymptotic result is based on some technical assumptions that are not always satisfied in practice.
- In particular, the parameters under the null model cannot lie on the boundary of the parameter space.
- This is a problem for testing variance components when the null hypothesis is they are equal to zero, as they are constrained to be positive (or positive definite).

Testing variance components

- Consider the model:

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

where $u_i \sim_{iid} N(0, \sigma_u^2)$ and $\epsilon_{ij} \sim_{iid} N(0, \sigma_\epsilon^2)$.

- Suppose we want to test whether the intercepts of the groups are significantly different from one another.
- This is equivalent to testing

$$H_0 : \sigma_u^2 = 0 \text{ versus } H_a : \sigma_u^2 > 0.$$

Testing variance components

- Under certain independence assumptions, the asymptotic distribution when H_0 is true implies that there is a 50% chance that $\hat{\sigma}_u^2 = 0$.
- This leads to the following approximate result:

$$-2 \log(LR(\mathbf{y})) \sim 0.5\chi_0^2 + 0.5\chi_1^2$$

where χ_0^2 is a point mass at 0.

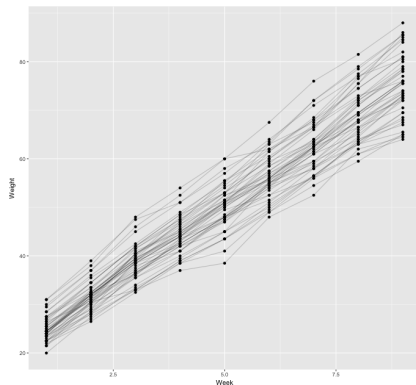
Testing variance components

- This results assumes that \mathbf{y} can be partitioned into subvectors that are independent.
- This assumption does not necessarily hold for all mixed models.
- The asymptotic distribution for general mixed models is more difficult.

Coding example

```
library(SemiPar)
library(lme4)
library(ggplot2)
data(pig.weights)

ggplot(pig.weights, aes(x = num.weeks, y = weight, group = id.num)) +
  geom_point() + geom_path(alpha = .2) +
  labs(x = "Week", y = "Weight")
```



Coding example

```
pig.mixed = lmer(weight ~ (1 | id.num) + num.weeks, data = pig.weights)
```

```
> summary(pig.mixed)
Linear mixed model fit by REML ['lmerMod']
Formula: weight ~ (1 | id.num) + num.weeks
Data: pig.weights
```

```
REML criterion at convergence: 2033.8
```

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-3.7390	-0.5456	0.0184	0.5122	3.9313

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
id.num	(Intercept)	15.142	3.891
Residual		4.395	2.096

```
Number of obs: 432, groups: id.num, 48
```

```
Fixed effects:
```

	Estimate	Std. Error	t value
(Intercept)	19.35561	0.60314	32.09
num.weeks	6.20990	0.03906	158.97

```
Correlation of Fixed Effects:
```

	(Intr)
num.weeks	-0.324

Coding example

```
pig.weights$pig.mixed.fit = fitted(pig.mixed)
ggplot(pig.weights, aes(x = num.weeks, y = weight, group = id.num)) +
  geom_point() + geom_path(alpha = .2) +
  labs(x = "Week", y = "Weight") +
  geom_{(1)}line(aes(y = pig.mixed.fit), color = "blue", alpha = .5)
```

