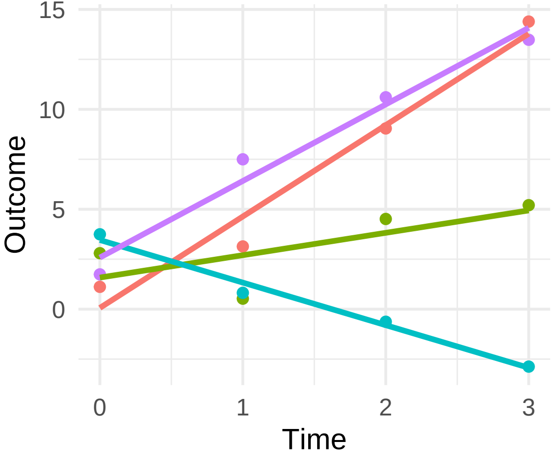
Linear Mixed Effects Models

What if we fit a *different* model for each individual?

We could find a unique β_{0i} and β_{1i} for each individual ...



Problems with this...

1. Generally these models will be **overspecified**.

Problems with this. . .

1. Generally these models will be **overspecified**.
2. This procedure would ignore **within-subject correlations**.

Alternative idea...

Share parameters.

Use **individual-level** terms in addition to **population-level** terms which are shared across the population.

Parameter Sharing in Practice

Instead of β_{0i} and β_{1i}, we could break these down into $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$.

Parameter Sharing in Practice

Instead of β_{0i} and β_{1i}, we could break these down into $\beta_{0i} = \beta_0 + b_{0i}$ and $\beta_{1i} = \beta_1 + b_{1i}$.

This way,

$$
\begin{aligned}
Y_{ij} &= \beta_{0i} + \beta_{1i} t_{ij} + \epsilon_{ij} \\
&= (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) t_{ij} + \epsilon_{ij} \\
&= \underbrace{(\beta_0 + \beta_1 t_{ij})}_{\text{Population Level}} + \underbrace{(b_{0i} + b_{1i} t_{ij})}_{\text{Individual Level}} + \epsilon_{ij}.
\end{aligned}
$$

Random Effects Terms

Consider One Individual

- Take Y_{ij} to be **one** individual's *continuous* outcome at time j.

Consider One Individual

- Take Y_{ij} to be **one** individual's *continuous* outcome at time j.
- We want to decompose this into three components:

Consider One Individual

- Take Y_{ij} to be **one** individual's *continuous* outcome at time j.
- We want to decompose this into three components:
  - The overall **population mean**: $X_{ij}'\beta$.

$$Y_{ij} = X_{ij}'\beta$$

Consider One Individual

- Take Y_{ij} to be **one** individual's *continuous* outcome at time j.
- We want to decompose this into three components:
  - The overall **population mean**: $X'_{ij}\beta$.
  - The **individual-level** mean: $Z'_{ij}b_i$.

$$Y_{ij} = X'_{ij}\beta + Z'_{ij}b_i$$

Consider One Individual

- Take Y_{ij} to be **one** individual's *continuous* outcome at time j.
- We want to decompose this into three components:
  - The overall **population mean**: $X_{ij}'\beta$.
  - The **individual-level** mean: $Z_{ij}'b_i$.
  - The measurement **variation** at time j: ϵ_{ij}.

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij}.$$

Regression Parameters with a Distribution

Recall that in **linear regression**, we take $\epsilon_i \sim N(0, \sigma^2)$ as a means of capturing the variation *across* the population.

Regression Parameters with a Distribution

Recall that in **linear regression**, we take $\epsilon_i \sim N(0, \sigma^2)$ as a means of capturing the variation *across* the population.

Can we do the same thing **here**?

Specification of a Linear Mixed Effects Model

Mathematical Notation for Mixed Effects Models

For a continuous variate Y_{ij}, we take

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij},$$

Mathematical Notation for Mixed Effects Models

For a continuous variate Y_{ij}, we take

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij},$$

where $b_i \sim N(0, D)$

Mathematical Notation for Mixed Effects Models

For a continuous variate Y_{ij}, we take

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij},$$

where $b_i \sim N(0, D)$, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ik_i})' \sim N(0, G_i)$,

Mathematical Notation for Mixed Effects Models

For a continuous variate Y_{ij}, we take

$$Y_{ij} = X_{ij}'\beta + Z_{ij}'b_i + \epsilon_{ij},$$

where $b_i \sim N(0, D)$, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ik_i})' \sim N(0, G_i)$, and with

$$b_i \perp \epsilon_i.$$

Mathematical Notation for Mixed Effects Models

For a continuous variate Y_{ij}, we take

$$Y_{ij} = X'_{ij}\beta + Z'_{ij}b_i + \epsilon_{ij},$$

where $b_i \sim N(0, D)$, $\epsilon_i = (\epsilon_{i1}, \cdots, \epsilon_{ik_i})' \sim N(0, G_i)$, and with

$$b_i \perp \epsilon_i.$$

Typically, we will set $G_i = \sigma^2 I$ to:

1. maintain the interpretation as **sampling error**; and
2. ensure **identifiability**.

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_i b_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_i b_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.
- If we consider the marginal distribution of Y_i we find:

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_ib_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.
- If we consider the marginal distribution of Y_i we find:
  - $E[Y_i] = X_i\beta$ since $E[b_i] = 0$.

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_i b_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.
- If we consider the marginal distribution of Y_i we find:
  - $E[Y_i] = X_i\beta$ since $E[b_i] = 0$.
  - $\text{var}(Y_i) = Z_i D Z_i' + G_i$, since $\text{var}(Z_i b_i) = Z_i \,\text{var}(b_i) Z_i' = Z_i D Z_i'$.

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_i b_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.
- If we consider the marginal distribution of Y_i we find:
  - $E[Y_i] = X_i\beta$ since $E[b_i] = 0$.
  - $\text{var}(Y_i) = Z_i D Z_i' + G_i$, since $\text{var}(Z_i b_i) = Z_i \,\text{var}(b_i) Z_i' = Z_i D Z_i'$.
  - Combining these results we find that, under this assumed model,

$$Y_i \sim N\left(X_i\beta, Z_i D Z_i' + G_i\right).$$

Mean, Variance, and Distribution

- If we condition on the **random effects** $E[Y_i|b_i] = X_i\beta + Z_i b_i$ and $\text{var}(Y_i|b_i) = \text{var}(\epsilon_i|b_i) = G_i$.
- If we consider the marginal distribution of Y_i we find:
  - $E[Y_i] = X_i\beta$ since $E[b_i] = 0$.
  - $\text{var}(Y_i) = Z_i D Z_i' + G_i$, since $\text{var}(Z_i b_i) = Z_i \text{var}(b_i) Z_i' = Z_i D Z_i'$.
  - Combining these results we find that, under this assumed model,

$$Y_i \sim N\left(X_i\beta, Z_i D Z_i' + G_i\right).$$

- This is a *specific* form of a **linear marginal model**!

Specific Examples: Random Intercept and Random Slope Models

The Random Intercept Model

The most basic version of a mixed effects model takes $Z_{ij} = 1$, and as such, b_i is a scalar for each individual.

This is called the **random intercept model**.

The Random Intercept Model

We have that

$$Y_i = \beta_0 + b_{0i} + \widetilde{X}_i\beta + \epsilon_i,$$

with $b_{0i} \sim N(0, \sigma_b^2)$ and $\epsilon_i \sim N(0, G_i = \sigma^2 I)$.

The Random Intercept Model

We have that

$$Y_i = \beta_0 + b_{0i} + \widetilde{X}_i\beta + \epsilon_i,$$

with $b_{0i} \sim N(0, \sigma_b^2)$ and $\epsilon_i \sim N(0, G_i = \sigma^2 I)$.

This gives

$$\text{cov}(Y_{ij}, Y_{il}) = \frac{\sigma_b^2}{\sigma_b^2 + \sigma^2}.$$

As a result, a random intercept model imposes the compound symmetry assumption!

The Random Intercept and Slope Model

If instead of *just* a random intercept, we also include a **random time slope** we get

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \widetilde{X}'_{ij}\beta + \epsilon_{ij}.$$

The Random Intercept and Slope Model

If instead of *just* a random intercept, we also include a **random time slope** we get

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 t_{ij} + b_{1i} t_{ij} + \widetilde{X}'_{ij}\beta + \epsilon_{ij}.$$

Here, D will be given by the variance of each b_{0i} and b_{1i}, as well as by the covariance between these terms.

The within-subject correlation will be time dependent in this model automatically!

Parameter Estimation and Hypothesis Testing

This is a Parametric Model

Recall that this model is parametric, we have assumed normality. As a result, we can estimate $\widehat{\beta}$ through maximum likelihood (or REML) directly!

This is a Parametric Model

Recall that this model is parametric, we have assumed normality. As a result, we can estimate $\widehat{\beta}$ through maximum likelihood (or REML) directly!

This will give the (familiar) asymptotic results where

$$\widehat{\beta} \quad \dot{\sim} \quad N\left(\beta, \left[\sum_{i=1}^{n} X_i' V_i^{-1}(\theta) X_i\right]^{-1}\right),$$

with $V_i(\theta) = \text{var}(Y_i) = Z_i D Z_i' + G_i$.

Hypothesis Testing

- Because this is a likelihood based methodology, we can use **LRTs**.

Hypothesis Testing

- Because this is a likelihood based methodology, we can use **LRTs**.
- The parameters for the covariances, denoted θ, will have some **regularity concerns**.

Hypothesis Testing

- Because this is a likelihood based methodology, we can use **LRTs**.
- The parameters for the covariances, denoted θ, will have some **regularity concerns**.
- Can use the standard **information criteria** as well!

Response Prediction (BLUPs)

Estimation versus Prediction

We saw **estimation** of the parameters β, but the b_i are random! As a result, we must **predict** them.

The best† predictor for b_i will be $E[b_i|Y_i]$, a quantity that we call the **best linear unbiased prediction** (or BLUP).

The BLUP

- The BLUP will take the form

$$E[b_i|Y_i] = DZ_i'V_i^{-1}(Y_i - X_i\beta).$$

The BLUP

- The BLUP will take the form

$$E[b_i|Y_i] = DZ_i' V_i^{-1}(Y_i - X_i\beta).$$

- This can be estimated using $\widehat{D}$, $\widehat{V}$, and $\widehat{\beta}$.

The BLUP

- The BLUP will take the form

$$E[b_i|Y_i] = DZ_i'V_i^{-1}(Y_i - X_i\beta).$$

- This can be estimated using $\widehat{D}$, $\widehat{V}$, and $\widehat{\beta}$.
- Once estimated, we can estimate outcomes as

$$\widehat{Y}_i = X_i\widehat{\beta} + Z_i\widehat{b}_i = \cdots = \widehat{G}_i\widehat{V}_i^{-1}X_i\beta + [I - \widehat{G}_iV_i^{-1}]Y_i.$$

The BLUP

- The BLUP will take the form

$$E[b_i | Y_i] = DZ_i' V_i^{-1}(Y_i - X_i\beta).$$

- This can be estimated using $\widehat{D}$, $\widehat{V}$, and $\widehat{\beta}$.
- Once estimated, we can estimate outcomes as

$$\widehat{Y}_i = X_i\widehat{\beta} + Z_i\widehat{b}_i = \cdots = \widehat{G}_i \widehat{V}_i^{-1} X_i\beta + [I - \widehat{G}_i V_i^{-1}]Y_i.$$

  - This is a **weighted average** between the estimated population mean $(X_i\beta)$ and the individual observation Y_i.

The BLUP

- The BLUP will take the form

$$E[b_i|Y_i] = DZ_i'V_i^{-1}(Y_i - X_i\beta).$$

- This can be estimated using $\widehat{D}$, $\widehat{V}$, and $\widehat{\beta}$.
- Once estimated, we can estimate outcomes as

$$\widehat{Y}_i = X_i\widehat{\beta} + Z_i\widehat{b}_i = \cdots = \widehat{G}_i\widehat{V}_i^{-1}X_i\beta + [I - \widehat{G}_iV_i^{-1}]Y_i.$$

  - This is a **weighted average** between the estimated population mean ($X_i\beta$) and the individual observation Y_i.
  - When G_i is **large** (more within-subject variation than between) there is more weight to the population average, and vice-versa.

Summary

- Marginal models are only able to estimate **population-level** effects.

Summary

- Marginal models are only able to estimate **population-level** effects.
- **Subject-level** effects are often relevant, and can be estimated using **mixed effects models**.

Summary

- Marginal models are only able to estimate **population-level** effects.
- **Subject-level** effects are often relevant, and can be estimated using **mixed effects models**.
- Linear mixed effects models are comprised of both **fixed effects** and **random effects** (which are individual specific).

Summary

- Marginal models are only able to estimate **population-level** effects.
- **Subject-level** effects are often relevant, and can be estimated using **mixed effects models**.
- Linear mixed effects models are comprised of both **fixed effects** and **random effects** (which are individual specific).
- We make normality assumptions, allowing for **standard asymptotic theory** when these are valid.

Summary

- Marginal models are only able to estimate **population-level** effects.
- **Subject-level** effects are often relevant, and can be estimated using **mixed effects models**.
- Linear mixed effects models are comprised of both **fixed effects** and **random effects** (which are individual specific).
- We make normality assumptions, allowing for **standard asymptotic theory** when these are valid.
- Two basic, common models (**random intercept** and **random intercept and slope**) capture correlation structures that we have previously seen.

Summary

- Marginal models are only able to estimate **population-level** effects.
- **Subject-level** effects are often relevant, and can be estimated using **mixed effects models**.
- Linear mixed effects models are comprised of both **fixed effects** and **random effects** (which are individual specific).
- We make normality assumptions, allowing for **standard asymptotic theory** when these are valid.
- Two basic, common models (**random intercept** and **random intercept and slope**) capture correlation structures that we have previously seen.
- We can use the **BLUP** to estimate individual effects, as-is necessary.

Drawbacks to Marginal Effects Models

Mixed effect models can be used to estimate **both** population-level and subject-level responses... what are the drawbacks?

Drawbacks to Marginal Effects Models

Mixed effect models can be used to estimate **both** population-level and subject-level responses... what are the drawbacks?

- They make distributional assumptions (GEEs did not).

Drawbacks to Marginal Effects Models

Mixed effect models can be used to estimate **both** population-level and subject-level responses. . . what are the drawbacks?

- They make distributional assumptions (GEEs did not).
- They may imply overly complex structures at the marginal level.