What are Marginal Models, and Why Do We Fit Them?

Brady T. West

Overview

So far, we've discussed the use of multilevel models with **random effects** to capture dependencies in clustered/longitudinal data sets

Overview

So far, we've discussed the use of multilevel models with **random effects** to capture dependencies in clustered/longitudinal data sets

Now, we shift focus to **marginal models** as alternative tools for capturing *within-cluster* correlations

Overview

So far, we've discussed the use of multilevel models with **random effects** to capture dependencies in clustered/longitudinal data sets

Now, we shift focus to **marginal models** as alternative tools for capturing *within-cluster* correlations

Key Distinction: In marginal models, we do ***not*** include random effects, because we are not interested in estimating between-cluster variance in coefficients

We still model **within-cluster** correlations

What are Marginal Models?

- General class of statistical models used to model dependent data, where observations within a randomly sampled cluster may be correlated
- We are interested in the estimation of overall, population-averaged relationships between independent variables (IVs) and dependent variables (DVs), across all clusters!

What are Marginal Models?

- General class of statistical models used to model dependent data, where observations within a randomly sampled cluster may be correlated
- We are interested in the estimation of overall, population-averaged relationships between independent variables (IVs) and dependent variables (DVs), across all clusters!

Reminder: In marginal models, we don't allow coefficients to randomly vary across clusters (a key feature of *multilevel* models)

What are Marginal Models?

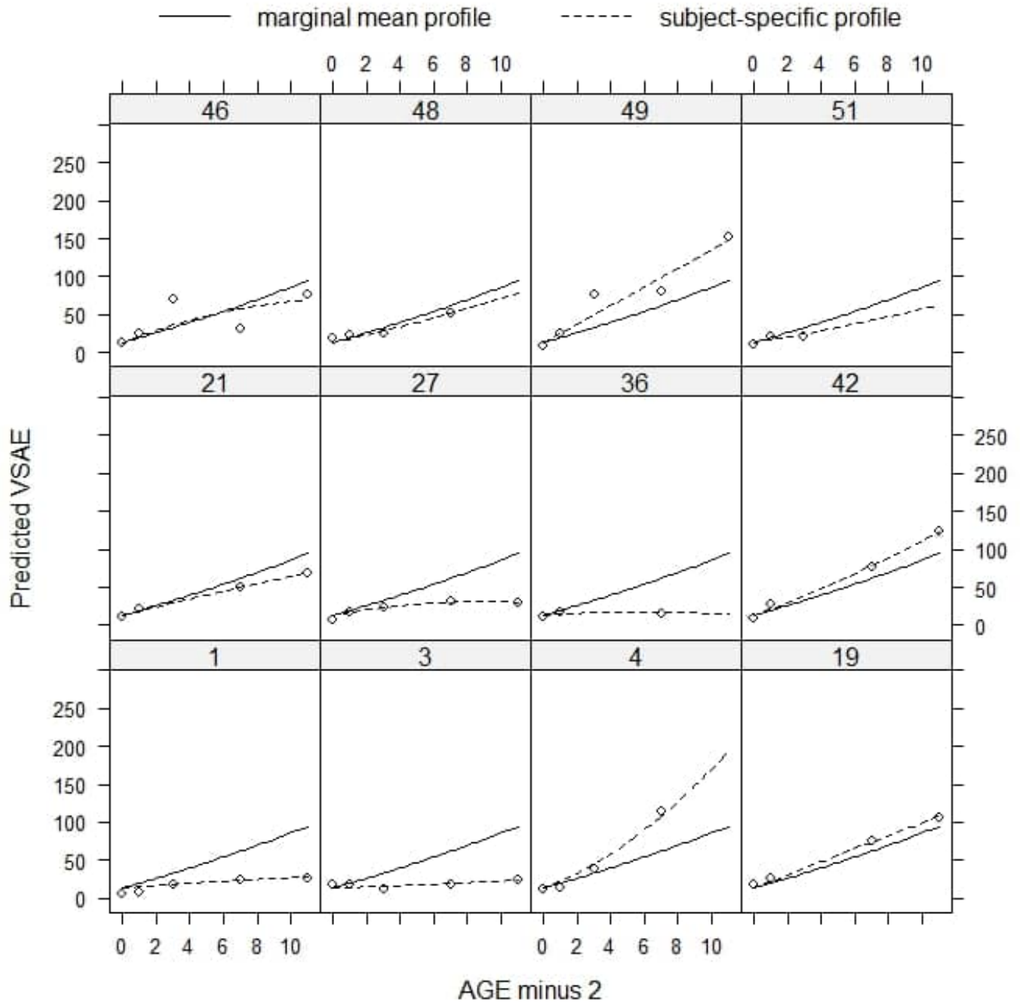
- General class of statistical models used to model dependent data, where observations within a randomly sampled cluster may be correlated
- We are interested in the estimation of overall, population-averaged relationships between independent variables (IVs) and dependent variables (DVs), across all clusters!

Reminder: In marginal models, we don't allow coefficients to randomly vary across clusters (a key feature of *multilevel* models)

Goal: Make inference about these overall, marginal relationships, with standard errors that reflect clustering

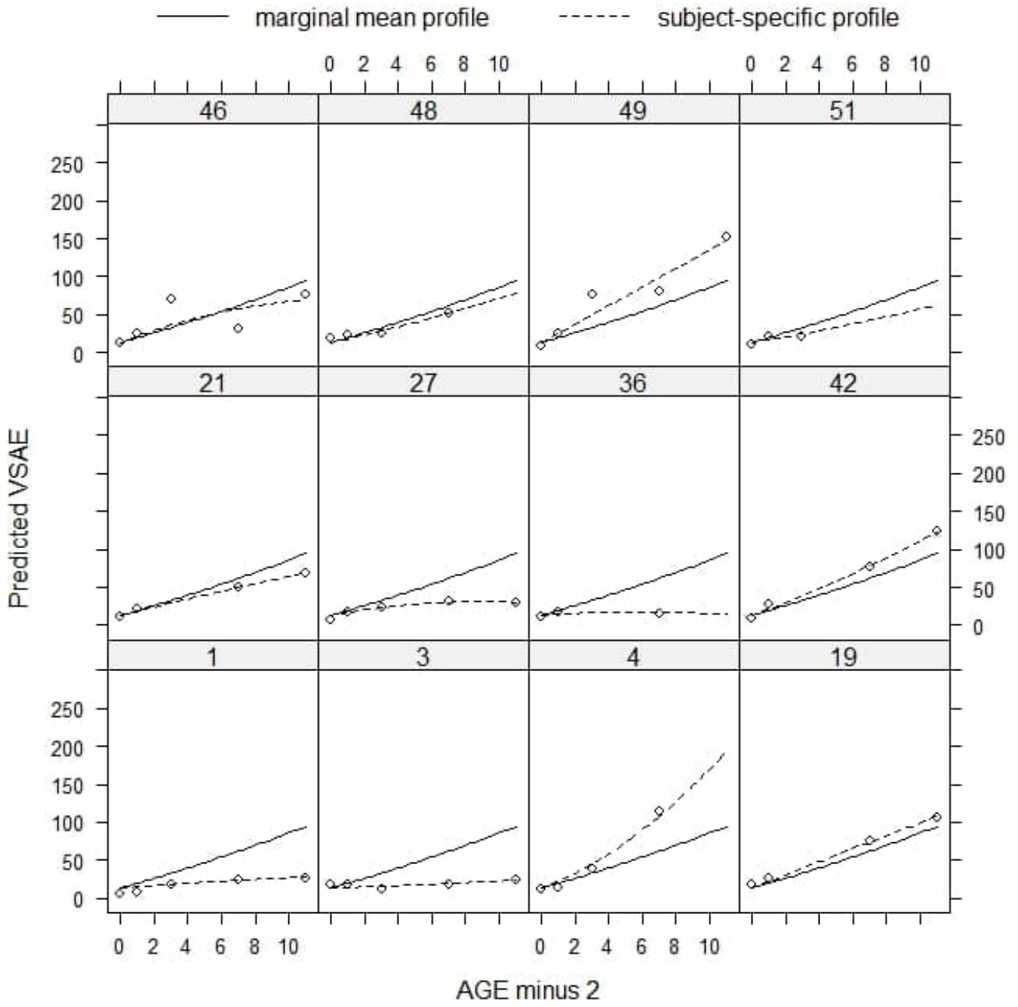
What are Marginal Models?

- Multilevel models capture dependencies by allowing coefficients to randomly vary across clusters



What are Marginal Models?

- Multilevel models capture dependencies by allowing coefficients to randomly vary across clusters
- Marginal models simply look at overall relationships, and make sure that standard errors reflect dependencies



Marginal Models

1. Explicitly select a structure for the mean of the dependent variable, usually defined by regression coefficients and predictor variables

(same as before!)

Marginal Models

1. Explicitly select a structure for the mean of the dependent variable, usually defined by regression coefficients and predictor variables

(same as before!)

2. Select structure that makes sense for the **variances and covariances** of observations coming from the same cluster that are not explained by the selected predictors

Marginal Models

1. Explicitly select a structure for the mean of the dependent variable, usually defined by regression coefficients and predictor variables

(same as before!)

2. Select structure that makes sense for the **variances <u>and</u> <u>covariances</u>** of observations coming from the same cluster that are <u>not explained</u> by the selected predictors

3. Compare the fits of models with different choices for this variance-covariance structure, and choose the best fit

Marginal Models

***Example**: Continuous longitudinal measurements on a dependent variable (DV) for subjects in a clinical study (e.g. headache pain)*

Marginal Models

***Example**: Continuous longitudinal measurements on a dependent variable (DV) for subjects in a clinical study (e.g. headache pain)*

After accounting for the relationships of time and experimental group, unexplained "errors" in the measures of the DV within a subject might follow an **auto-regressive** covariance structure:

- Errors close to each other in time have a stronger correlation than errors farther apart
- Observations have constant variance over time

Marginal Models

***Example**: Continuous longitudinal measurements on a dependent variable (DV) for subjects in a clinical study (e.g. headache pain)*

After accounting for the relationships of time and experimental group, unexplained "errors" in the measures of the DV within a subject might follow an **auto-regressive** covariance structure:

- Errors close to each other in time have a stronger correlation than errors farther apart
- Observations have constant variance over time

For clustered data, we might assume that the errors within a cluster follow an **exchangeable** covariance structure (constant variance and covariance)

When Can We Fit Marginal Models?

Rather than assuming i.i.d. data!)

- ☑ We have data from a longitudinal or clustered study design that **introduces dependencies in the collected data**, and we need to model those dependencies to obtain **accurate inferences** about relationships of interest

When Can We Fit Marginal Models?

Rather than assuming i.i.d. data!)

- ☑ We have data from a longitudinal or clustered study design that **introduces dependencies in the collected data**, and we need to model those dependencies to obtain **accurate inferences** about relationships of interest
- ☑ We have **no interest** in estimating between-subject or between-cluster variance in the coefficients of interest

When Can We Fit Marginal Models?

Rather than assuming i.i.d. data!)

- ☑ We have data from a longitudinal or clustered study design that **introduces dependencies in the collected data**, and we need to model those dependencies to obtain **accurate inferences** about relationships of interest
- ☑ We have **no interest** in estimating between-subject or between-cluster variance in the coefficients of interest
- ☑ We wish to make inference about **overall, marginal relationships between IVs and DVs** in the target population; that is, we do not wish to condition on random effects of subjects or clusters in the modeling

Why do we Fit Marginal Models?

- These models offer some advantages over other approaches for **dependent data** (e.g., multilevel modeling):
  - Quicker computational times; faster estimation
  - Robust standard errors that reflect the specified correlation structure
  - Easier accommodation of non-normal outcomes (recall that multilevel models for non-normal outcomes can take a while to estimate!)

Why do we Fit Marginal Models?

- These models offer some advantages over other approaches for **dependent data** (e.g., multilevel modeling):
  - Quicker computational times; faster estimation
  - Robust standard errors that reflect the specified correlation structure
  - Easier accommodation of non-normal outcomes (recall that multilevel models for non-normal outcomes can take a while to estimate!)

Remember: We can no longer make inference about between-cluster variance in the coefficients of interest!

What's Next?

- We'll revisit earlier examples of multilevel modeling, and consider alternative marginal modeling approaches
- We'll discuss how to estimate these models, and methods for assessing model fit
- Finally, we will present examples, and you'll be able to walk through examples of fitting these models using Python!