LINEAR MODELS FOR LONGITUDINAL DATA

Notation and Overview and Simple Examples

- Simple marginal (population average) **mean model** specification and interpretation
- Simple **variance-covariance-correlation model** examples

Notation

- The general model for subject i is:

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i \tag{1}$$

where

$$\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$$

$$X_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})^T$$

$$\boldsymbol{x}_{ij}^T = (1, x_{ij1}, \ldots, x_{ijp})$$

$$\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$$

What goes into the predictor vector $\boldsymbol{x}_{ij}$ will vary widely from context to context (but will generally include time, baseline covariates, functions thereof,...)

- For the vector $\boldsymbol{\epsilon}_i$ of error terms, assume

$$\boldsymbol{\epsilon}_i \sim MVN(0, V_i) \ , \quad \boldsymbol{\epsilon}_i \amalg X_i \ , \tag{2}$$

($\amalg$ means "independent of") where V_i is the **variance-covariance** matrix of $\boldsymbol{\epsilon}_i$

- Note: The **multivariate normality** assumption will sometimes be important and other times not be so critical, depending on the analysis

  For now, take multivariate normality as a **working model**, with the critical components being

$$\mathrm{E}(\boldsymbol{\epsilon}_i) = 0 \quad \text{and} \quad \mathrm{var}(\boldsymbol{\epsilon}) = V_i \tag{3}$$

- V_i is a matrix of dimension $n_i \times n_i$

$$\mathbf{V_i} = \begin{pmatrix} v_{i11} & v_{i12} & \cdots \\ v_{i21} & v_{i22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

with diagonal elements

$$v_{ijj} = \text{var}(\epsilon_{ij}) = \sigma_{ij}^2$$

and off-diagonal elements

$$v_{ijk} = \text{cov}(\epsilon_{ij}, \epsilon_{ik}) = \sigma_{ij}\sigma_{ik}\rho_{ijk} \ ,$$

where

– $\sigma_{ij}^2 = \text{var}(\epsilon_{ij})$

– $\rho_{ijk} = \text{corr}(\epsilon_{ij}, \epsilon_{ik})$

– $\sigma_{ij}\sigma_{ik}\rho_{ijk} = v_{ijk} = \text{cov}(\epsilon_{ij}, \epsilon_{ik})$

- We may sometimes write σ^2 without the subscripts i or j and ρ_{jk} without the subscript i, for simplicity
- Often we will place **model structure** on V_i (examples later)

- Model (1), with (2) or (3), form a **marginal model** with the two components:
  - A mean model: $\mathrm{E}(\boldsymbol{Y}_i|X_i) = X_i\boldsymbol{\beta}$, expressing the **trend** over time or other variation in Y_{ij} with respect to predictors
  - A variance-covariance model: $\mathrm{var}(\boldsymbol{Y}_i|X_i) = \mathrm{var}(\boldsymbol{\epsilon}_i) = V_i$, capturing variability and covariability in components of $\boldsymbol{Y}_i$ not explained by X_i
- Important note: in a basic **marginal model**, the β-coefficients take the same interpretation as in an ordinary linear regression model for cross-sectional data (as if all Y_{ij}'s were independent)
- In this lecture, we will consider $\mathrm{E}(\boldsymbol{Y}_i|X_i)$ to be a basic marginal model, and then in future lectures, we will extend this to incorporate:
  - time profile analysis, perhaps with key covariates of interest
  - longitudinal models for time-varying covariates, incorporating within- and between-subject terms
  - variations and combinations
- We will also consider various **models** for the variance σ^2_{ij} and the correlation ρ_{ijk} or covariance, including
  - **Unstructured** correlation models: Allowing the correlation to be different for

each pair of time points

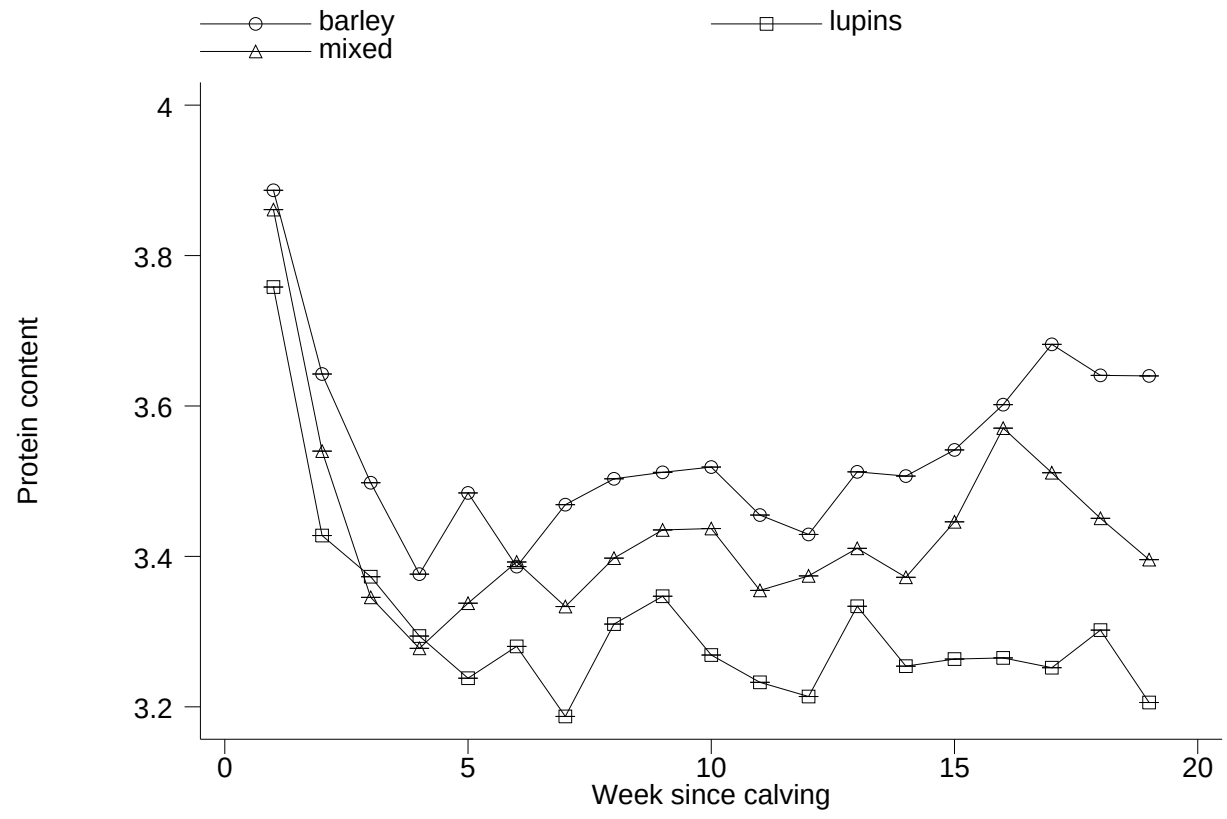
– **Correlation structure** or **correlation pattern** models: Modeling the correlation as a function of the two time points t_{ij} and t_{ik}

– **Random effects models**: Modeling each subject's trajectory with random intercepts, slopes, etc., specific to that subject . . . here, the covariance structure will arise implicitly

- In addition, modelling the variance-covariance might be completely avoided in a **fixed effects** model
- **Note:** Sometimes you see $\mathrm{E}(\boldsymbol{Y}_i)$ and $\mathrm{var}(\boldsymbol{Y}_i)$, but conditioning on X_i is a more precise description of the actual model
- We will generally assume that subjects are **independent**, so we do not need to be concerned with correlation among observations on different subjects

Mean Model Example

- **Example:** Protein content in cows' milk depends on diet
- Scientific goal: Relationship of time profile (since calving) of protein content of milk to diet (lupins, barley, mixed)
- Examine mean response profile for each diet group



- Preliminary observations:
  - Protein content drops strongly over first several weeks after calving and then levels off
  - Cows on diets with higher barley content have higher protein content (on average)
  - Mean response profiles for each group follow a somewhat similar pattern over time
- Linear model for each dietary group:
- What does the model look like and what are the key parameters in this model?

$$\begin{aligned} \mathrm{E}(Y_{ij}|X_{ij}) &= \beta_0 + \beta_1(\mathsf{week}_{ij} - 1) + \beta_2 I(\mathsf{diet}_i = 2) + \beta_3 I(\mathsf{diet}_i = 3) \\ &\quad + \beta_4 I(\mathsf{diet}_i = 2) \times (\mathsf{week}_{ij} - 1) \\ &\quad + \beta_5 I(\mathsf{diet}_i = 3) \times (\mathsf{week}_{ij} - 1) \ , \end{aligned}$$

where Y_{ij} is protein content for ith cow on jth time point

Then

- $\beta_1 =$ difference in mean response for each unit difference in time (time slope) among cows on diet 1

- $\beta_2 =$ difference in mean response comparing cows on diet 2 to those in diet 1, **at** week 1 (baseline treatment effect)
- $\beta_4 =$ difference in time slope of mean response comparing diet 2 to diet 1 (response drops more quickly on diet 2 than diet 1)

  or

  estimated difference in treatment effect comparing diet 2 to diet 1 for each unit difference in time (slope of treatment effect; treatment effect greater with increasing time)

- Note that interpretation of β's is just as it is in OLS
- This is an example of a **profile analysis model**, which is often employed when comparing trajectories across treatment groups. It is probably not a very good model for these data (linearity assumption not appropriate); we will explore ways to improve it later.

- **Example:** Arm circumference in Nepalese children is related to weight, with potential confounders age and sex
- Linear regression model of arm circumference on weight, age and sex:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \epsilon_{ij}$$

  where
  - $Y_{ij} =$ arm circumference
  - $x_{ij1} =$ weight
  - $x_{ij2} =$ age
  - $x_{ij3} = x_{i\cdot 3} =$ sex
- **Marginal model** interpretation of β_1 (coefficient of weight):
  Estimated difference in mean arm circumference comparing two children (actually, two randomly-selected children & time-points from population) differing in weight by 1 kg, adjusting for age and sex

- This is the **same interepretation** as β_1 with **cross-sectional** data (marginal model interpretation)
- This is a simple example of models for the mean response $\mathrm{E}(Y_{ij}|X_{ij})$. We need to complement our **mean model** with a **variance-covariance-correlation model**, in order to do estimation and inference

Simple Variance-Covariance-Correlation Model Examples
Equal Variance and Exchangeable Correlation

- Assume $\rho_{jk} = \rho$ (same for all pairs of observations on a typical subject) and also that $\mathrm{var}(Y_{ij})$ is constant over j
- The variance-covariance matrix V_0 (the variance-covariance matrix for a "typical" subject) for five observations on a subject in the Nepal data is then

$$V_0 = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho & \rho \\ \rho & \rho & 1 & \rho & \rho \\ \rho & \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & \rho & 1 \end{pmatrix}$$

- This is called the **uniform**, **exchangeable** or **compound symmetry** correlation model

- In the Nepal data, the **within-subject** correlation coefficient is about $0.75 = \widehat{\rho}$. (we will see how we get this)
  - the correlation between two arm circumference measures on the same child is about 0.75
  - **the foregoing mean model** removes any similarities in arm circumference due to weight, age or sex. i.e., $\hat{\rho}$ is **adjusted** for weight, age and sex
  - the correlation is **adjusted for the mean model**, i.e., the correlation applies to what is left over after removing trend due to $\boldsymbol{x}_{ij}$
- In this model, the variance σ^2_{ij} is constant over time.
  - Most non-random effects models we will examine have **constant variance**
- Models with variance σ^2_{ij} **changing over time**:
  - unstructured variance-covariance model
  - random effects models with random slopes

- How does the exchangeable correlation model arise? Suppose that

$$\epsilon_{ij} = U_i + Z_{ij}$$

where $U_i \sim N(0, \nu^2)$ and is constant over time within subject, $Z_{ij} \sim N(0, \tau^2)$ and varies over time within subject, and the U_i's and Z_{ij}'s are all independent of one another

- Then:

$$\mathrm{var}(\epsilon_{ij}) = \mathrm{var}(U_i + Z_{ij}) = \mathrm{var}(U_i) + \mathrm{var}(Z_{ij}) = \nu^2 + \tau^2$$

$$\begin{aligned} \mathrm{cov}(\epsilon_{ij}, \epsilon_{ik}) &= \mathrm{E}(\epsilon_{ij}\epsilon_{ik}) \\ &= E[(U_i + Z_{ij})(U_i + Z_{ik})] = E(U_i^2) \\ &= \nu^2 \end{aligned}$$

$$\rho_{jk} = \rho = \frac{\mathrm{cov}(\epsilon_{ij}, \epsilon_{ik})}{\mathrm{var}(\epsilon_{ij})} = \frac{\nu^2}{\nu^2 + \tau^2}$$

(same for all pairs of time points)

- The U_i allows each child to have his/her own intercept, given by

$$\beta_0 + U_i$$

the term U_i is called a **random effect**; random effects are one way to generate v-c-c models

- In this model, estimates are approximately

$$\widehat{\text{var}}(U_i) = \hat{\nu}^2 = 0.41$$

and

$$\widehat{\text{var}}(Z_{ij}) = \hat{\tau}^2 = 0.13$$

and

$$\widehat{\rho} = \frac{.41}{.41 + .13} \approx 0.76$$

- This is the same set-up as the variance structure model we introduced earlier in our exploratory techniques:
  - $\nu^2 =$ between-subject variance
  - $\tau^2 =$ within-subject variance

- More details about estimation in coming lectures
- This **mean** and **variance-covariance-correlation** model can be fitted in SAS:

```
proc mixed data=nepal method=ML;
class id sex;
model arm=wt age sex/ s;
random intercept/ subject=id;
run;
```

- Notes on the syntax:

  `proc mixed`: Fit a linear mixed-effects model

  `random intercept`: add a random intercept U_i

  `method=ML`: estimate the model using maximum likelihood

- In R: can use `lmer()` function in `lme4` package

- Part of results:

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	id	0.4106
Residual		0.1275

Solution for Fixed Effects

Effect	sex		Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept			8.9626	0.1633	195	54.89	<.0001
wt			0.6630	0.02326	678	28.51	<.0001
age			-0.05949	0.003620	678	-16.43	<.0001
sex		1	-0.2920	0.09528	678	-3.06	0.0023
sex		2	0	.	.	.	.

- Interpretation of $\hat{\beta}_1 = 0.663$ is the same as in an ordinary linear regression model
- The variance parameter estimates are

$$\widehat{\text{var}}(U_i) = \hat{\nu}^2 = 0.4106$$

and

$$\widehat{\text{var}}(Z_{ij}) = \hat{\tau}^2 = 0.1275$$

- The estimated **within-subject** correlation coefficient is therefore

$$\widehat{\rho} = \frac{.4106}{.4106 + .1275} = \frac{\text{between-subj variance}}{\text{total variance}} = 0.763$$

Equal Variance and Exponential Correlation

- A different model is to assume that the correlation of observations closer together in time is larger than that of observations farther apart
- One model for this is the **exponential** or **AR (1)** correlation model

$$\rho_{jk} = \alpha^{|t_j - t_k|}$$

- In the data on Nepalese children, if we let t indicate a four-month interval, then $t_1 = 1$, $t_2 = 2$, etc., and

$$\alpha = \rho_{j,j-1} = \text{correlation between observations } j \text{ \& } j-1$$

$$\alpha^2 = \rho_{j,j-2} = \text{correlation between observations } j \text{ \& } j-2$$

etc.

- So, the covariance matrix V_i for the five observations each one unit apart on a subject i is

$$\mathbf{V_0} = \sigma^2 \begin{pmatrix} 1 & \alpha & \alpha^2 & \alpha^3 & \alpha^4 \\ \alpha & 1 & \alpha & \alpha^2 & \alpha^3 \\ \alpha^2 & \alpha & 1 & \alpha & \alpha^2 \\ \alpha^3 & \alpha^2 & \alpha & 1 & \alpha \\ \alpha^4 & \alpha^3 & \alpha^2 & \alpha & 1 \end{pmatrix}$$

- In these data, the estimated **within-subject** correlation coefficient for two observations separated by 4 months is approximately

$$\widehat{\text{corr}}(\epsilon_{ij}, \epsilon_{i,j-1}) = \hat{\alpha} = 0.80$$

(again, adjusted for weight, age and sex)

- How does this model arise? Suppose that $\epsilon_{ij} = W_{ij}$, and

$$W_{i1} \sim N(0, \delta^2)$$

$$W_{ij} = \alpha W_{i,j-1} + E_{ij}, \quad j = 2, 3, \ldots$$

  where $E_{ij} \sim N\{(0, \delta^2(1-\alpha^2)\}$ are all independent of one another
- This is sometimes called a **first order autoregressive** (AR(1)) model
- W_{ij} is a **first-order autoregressive process** with

$$\begin{aligned} \text{var}(\epsilon_{ij}) &= \text{var}(W_{ij}) = \text{var}(\alpha W_{i,j-1} + E_{ij}) \\ &= \alpha^2 \text{var}(W_{i,j-1}) + \text{var}(E_{ij}) \\ &= \alpha^2\delta^2 + \delta^2(1-\alpha^2) = \delta^2 \end{aligned}$$

$$\text{cov}(\epsilon_{ij}, \epsilon_{i,j-1}) = E[(\alpha W_{i,j-1} + E_{ij})W_{i,j-1}] = \alpha \text{var}(W_{i,j-1}) = \alpha\delta^2$$

$$\rho_{j,j-1} = \frac{\text{cov}(\epsilon_{ij}, \epsilon_{i,j-1})}{\text{var}(\epsilon_{ij})} = \alpha$$

$\rho_{j,j-2} = \alpha^2$ (work out at home!)

- This model allows each subject's error term ϵ_{ij} at a given time to be a function of his/her error term $\epsilon_{i.j-1}$ at the **previous** time
- This is a type of **transition** model for discrete times
- More generally, we could write $W_{ij} = W_i(t_{ij})$ where $W_i(\cdot)$ is a random Gaussian process with:
  - $\mathrm{E}\{W_{ij}(\cdot)\} = 0$,
  - $\mathrm{var}\{W_i(\cdot)\} = \delta^2$,
  - $\mathrm{corr}\{W_i(t), W_i(t+u)\} = \alpha^u, u > 0$

Unstructured Correlation Model

- Suppose that:
  - We **did not** want to make **any assumptions** about the correlation of the repeated observations on an individual, not even stationarity
  - We had approximately **balanced** data
- Then, we might fit a model with no structure on the correlation matrix (i.e., an **unstructured model**)
- To do this, we need a finite number of time points (nearly balanced data), because the correlation between each two time points will be estimated separately
- We could fit such a model with the Nepal data, using `obs` as the time variable (Why wouldn't age work?)
- This model could be generalized even further to **allow for different variances at each time point**

Important Themes in This and Coming Lectures

- These are **marginal models**: The interpretation of the regression coefficients is the **same** as that in cross-sectional data
- Accounting for the association / correlation in LD is important for correct inferences on regression coefficients (β)
  - statistical efficiency
  - correct standard errors
- Association can be included via:
  - structured correlation models (exchangeable, exponential, . . .)
  - unstructured correlation models (for balanced data)
  - random effects model (e.g., random intercept, could add slope)
  - fixed effects models
- Correlation can arise by way of **subject-specific** models (U_i) and/or **transition** models (W_{ij})
  - Exchangeable correlation model: subject-specific formulation
  - Exponential correlation model: transition model formulation

- Incorporating correlation into estimation of regression models is achieved via **Weighted Least Squares**, coming up
- V-C-C model parameters will be estimated via **Restriced Maximum Likelihood Estimation**, also coming up
- Correlation models can be **approximate**
  - We will call these **working correlation models** ("our best shot")
  - Regression coefficients estimates will still be correct
  - We will see how to "fix up" standard errors to account for inaccuracies in correlation models