A. GETTING STARTED

Learning Outcomes

By the end of today's class, you will be able to

I. Describe multilevel models as extension of the OLS model
II. Develop the framework for parameter inference in the MLM context
III. Implement the MLM model with real data in Python and interpret results

Where we're headed: Let's take a look at a visualization that will help motivate the new direction for this module: Hierachical Models[1]

Motivating Context: The 2023 Employment Insurance Coverage Survey (CEIS) was commissioned by the Labour Statistics Division (Statistics Canada) with the goal of tracking "the performance of the Employment Insurance (EI) program, by finding out how many people are covered by EI, what proportion of people receive benefits and which **groups** of people who may need EI do not get access to Employment Insurance"[2]. As a statistician on this project, you are tasked to use the CEIS data to investigate whether **HOURS** (**number of insurable hours** the year before respondent stopped working) is predictive of **LOOK_HRS** (**hours spent looking** for work per week). Observation units are individuals selected from the Labour Force Survey (LFS).

B. RECAP: Ordinary Least Squares

Recall from last week's lecture the simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \dots, n$$

where $\epsilon_i \sim N(0, \sigma^2)$ and $E(y_i) = \mu_i$.

Use matrix notation to represent this model

$\vec{y} =$	$\vec{\beta} =$	$X =$	$\vec{\epsilon} =$

thus

We showed the least squares estimate is equivalent to the maximum likelihood estimate in this context, with $\hat{\hat{\mu}} =$ $= H\vec{y}$. This is the Best Linear Unbiased Estimator (BLUE) with $V(\hat{\hat{\mu}}) = \sigma^2 H$.
Applying OLS to our CEIS data yields a significant linear relationship (p=0.001)

LOOK_HRS** = 8.1108 + 0.0013 ***HOURS

[1] Shout out to Dr. Michael Freeman at the at the University of Washington Information School. You can find the code here: https://github.com/mkfreeman/hierarchical-models/

[2] Retrieved from Odesi: https://odesi.ca/fr/d%C3%A9tails?id=/odesi/doi__10-5683_SP3_MYZ1UV.xml

C. CORE TOPIC: The Random Intercept Model

In some instances, it is known or expected that the relationship between the response and explanatory variables could be different for subgroups of the data.

Discussion: ***How could we build into our model the flexibility to effectively model the relationship between our explanatory and response differently for a sample of subgroups?***

"Once groupings are established, even if their establishment is effectively random, they will tend to become differentiated, and this differentiation implies that the group' and its members both influence and are influenced by the group membership. To ignore this relationship risks overlooking the importance of group effects, and may also render invalid many of the traditional statistical analysis techniques used for studying data relationships." **– Goldstein, 1999**

Consider the following generalization of the OLS model:

Level 1: $y_{ji} = \beta_{0j} + \beta_1 x_{ji} + \epsilon_{ji},\ j = 1, \ldots, J; i = 1, \ldots, n_j$

where

Level 2: $\beta_{0j} = \beta_0 + u_j$

represents a ***random intercept***, an explicit way to model variation between the J groups in the relationship between our explanatory variable and response. The model can be written without explicit reference to level as follows:

$$y_{ji} = \beta_0 + \beta_1 x_{ji} + u_j + \epsilon_{ji},\ \ j = 1, \ldots, J;\ i = 1, \ldots, n_j, n = \sum_{j=1}^{J} n_j$$

Example 1: For the 2023 Employment Insurance Coverage Survey, we would like to incorporate a two-level structure, where individuals are grouped according to a random sample of `INDUSTRY` (Ag/Forestry = 1, Ed/Health/Social Svcs = 5, Other = 6). Suppose there are 2 in category 1, 3 in category 2 and 2 in category 3 in our sample. For this specific example, represent the model in matrix form.

Sizes	**Individuals**	**Matrix Notation**
$J =$	$y\_\_\_\_ =$	
	$y\_\_\_\_ =$	
$n_1 =$	$y\_\_\_\_ =$	
$n_5 =$	$y\_\_\_\_ =$	
$n_6 =$	$y\_\_\_\_ =$	
	$y\_\_\_\_ =$	
$n =$	$y\_\_\_\_ =$	

More generally, we can express the model in matrix notation as

It is important to note that like $\vec{\epsilon}$, the vector $\vec{u}$ enters the model as a _______________ effect, whereas $\vec{\beta}$ is a vector of ________________ effects. Thus the model specification is not complete without distributional assumptions on the random components. Represent the following symbolically

We will assume independence

I. Between group effects and individual effects:
II. Between one group and another:

We will also assume that all effects are normally distributed:

$$\epsilon_{ji} \sim N(0, \sigma^2)$$

$$u_j \sim N(0, \tau^2)$$

This leads to the following formulation:

$$\vec{y} = X\vec{\beta} + Z\vec{u} + \vec{\epsilon}$$

$$\vec{u} \sim N(0, \tau^2 I), \qquad \vec{\epsilon} \sim N(0, \sigma^2 I)$$

We can use properties of random vectors to determine how to proceed with estimation in this context

If $\vec{w} = A\vec{v}$ then $E(\vec{w}) =$ and $V(\vec{w}) =$

and so

$$E(\vec{y}) = X\vec{\beta} \text{ and } V(\vec{y}) = V =$$

Example 2: In Example 2, the covariance matrix of the observations would end up being

$$V(\vec{y}) = V = \tau^2 ZZ^T + \sigma^2 I =$$

It follows from independence that $\vec{y} \sim N(X\vec{\beta}, V)$, thus

$$f_{\vec{y}}(\vec{y}) = ((2\pi)^n |V|)^{-1/2} exp\left(-\frac{1}{2}\left(\vec{y} - X\vec{\beta}\right)^T V^{-1}(\vec{y} - X\vec{\beta})\right), \vec{y} \in \mathbb{R}^n$$

Thus our log-likelihood becomes

$$l(\vec{\beta}, \sigma^2, \tau^2) =$$

and analogous to OLS, the maximum likelihood criterion corresponds to minimizing the quadratic form that appears in the argument of the exponential function in the multivariate normal distribution, i.e.

$$\hat{\vec{\beta}} = argmin\ ((\vec{y} - X\vec{\beta})^T V^{-1}(\vec{y} - X\vec{\beta}))$$

Discussion: ***What are some similarities and differences between the optimization step in OLS versus what we are facing here?***

D. THE ALGORITHM: Restricted Maximum Likelihood (REML)

We're about to jump in the deep end. We need some linear algebra results to push this forward

Key Linear Algebra and Multivariate Normal Results
(R1) $\frac{\partial}{\partial \vec{w}}(\vec{w}^T A \vec{w}) = 2A^T\vec{w}$
(R2) If $\begin{bmatrix}\vec{a}\\ \vec{b}\end{bmatrix} \sim N\left(\begin{bmatrix}\vec{\mu}_a\\ \vec{\mu}_b\end{bmatrix}, \begin{bmatrix}\Sigma_a & \Sigma_{ab}\\ \Sigma_{ba} & \Sigma_b\end{bmatrix}\right)$ then $E(\vec{a}\|\vec{b}) = \vec{\mu}_a + \Sigma_{ab}\Sigma_b^{-1}(\vec{b} - \vec{\mu}_b)$

When we determine our MLE for $\vec{\beta}$ in the OLS case, it's a dance that has two steps…

In the present case, because of the adaptivity problem, we use an iterative algorithm known as REML to estimate the unknown parameters.

STEP 1: Set initial values for error variance estimates (for example by method of moments.

STEP 2: Use (R1) to show $\frac{\partial}{\partial \vec{\beta}} l\left(\vec{\beta}, \sigma^2, \tau^2\right) =$

STEP 3: Estimate $\vec{\beta}$

STEP 4: Predict $\vec{u}$ using (R2)

STEP 5: Profile out fixed effects $\vec{\beta}$ and minimize $l\left(\vec{\beta}, \sigma^2, \tau^2\right)$ over (σ^2, τ^2). *Fisher Scoring* and *Average Information* are Newton-Raphson-type algorithms typically used in this context.

STEP 6: Check for convergence

Next Steps:

Walk through the Theory	**Apply the Model to CEIS Data**
https://bit.ly/WTTT_RI	https://bit.ly/CEIS_RI