Lecture 21
Multi-level GLM

General linear model

- The **general linear model (GLM)** approach treats the data as a linear combination of model functions (predictors) plus noise (error).

- The model functions are assumed to have *known* shapes (i.e., straight line, or known curve), but their amplitudes (i.e., slopes) are *unknown* and need to be estimated.

- The GLM framework encompasses many of the commonly used techniques in fMRI data analysis (and data analysis more generally).

- Analysis
  - Regression
  - One-sample/Two-sample t-test
  - Analysis of variance (ANOVA)
  - Analysis of covariance (ANCOVA)
  - Multivariate analysis of variance (MANOVA)
  - …

Slide credit: Tor Wager

GLM family

Closed form solution

- Simple regression
- ANOVA
- Multiple regression
- General Linear Model
  - Mixed effects/hierarchical
  - Timeseries models (e.g., autoregressive)
  - Robust, penalized regression (LASSO, Ridge)

Iterative solutions

- Generalized Linear Model
  - Non-normal errors
  - Binary, categorical outcomes (logistic regression)

Slide credit: Tor Wager

General linear model

Specify model → Estimate → Statistical inference → Scientific interpretation

In case of the simple regression,

1. Simplification : linear relationship
2. Find slope and intercept
3. Test slope: finding p-value
4. What is the meaning of relationship?

$$y = ax + b + e$$

Outcome = slope x predictor + intercept + error
(DV) (constant) (residual)

Slide credit: Tor Wager

Multiple regression

- Structural model for regression
  - Variables
  - Parameters

$$\underbrace{y_i}_{\text{Dependent variable}} = \underbrace{\beta_0}_{\text{intercept}} + \beta_1 \underbrace{x_{i1}}_{\text{predictor}} + \underbrace{\beta_2}_{\text{slope}} \underbrace{x_{i2}}_{\text{predictor}} + \ldots + \underbrace{\beta_k}_{\text{slope}} \underbrace{x_{ik}}_{\text{predictor}} + \underbrace{\varepsilon_i}_{\text{error}}$$

- Solve for beta vector which minimizes sum of squared residuals
  - Estimating $\widehat{\beta_n}$

- Matrix notation

$$y = X\beta + \varepsilon$$

Slide credit: Tor Wager

Multiple regression

- With matrix notation

$$Y = X\beta + \varepsilon$$

as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Observed data
Outcome data

Design matrix

Model parameters

Residuals

Slide credit: Tor Wager

Multiple regression

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Observed data
Outcome data

Design matrix

Model parameters

Residuals

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \beta_0 \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{1p} \end{bmatrix} + \cdots + \beta_p \begin{bmatrix} X_{1p} \\ X_{2p} \\ \vdots \\ X_{np} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min}\, S(\boldsymbol{\beta})$$

Objective function S:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} X_{ij}\beta_j \right|^2 = \left\| \mathbf{y} - \mathrm{X}\boldsymbol{\beta} \right\|^2$$

Ordinary least square solution

$$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^{\mathsf{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathsf{T}}\mathbf{y}.$$

https://en.wikipedia.org/wiki/Ordinary_least_squares

Slide credit: Tor Wager

A simple example:

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1p} \\ 1 & X_{21} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & X_{n1} & \dots & X_{np} \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Observed data
Outcome data

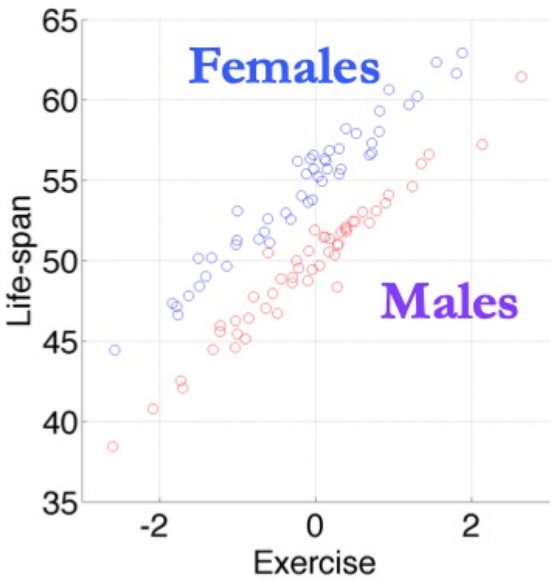
Design matrix

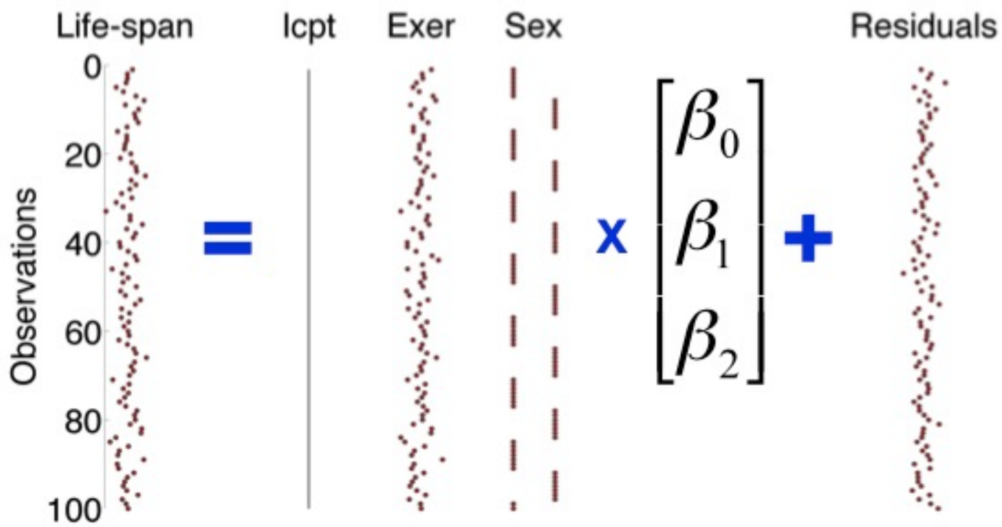
Model parameters

Residuals

- Does exercise predict life-span? (not real data)
- Control for other variables that might be important, e.g., male vs. female

$$y = ax + b + e$$





Slide credit: Tor Wager

Multi-level analysis

- What is the multilevel model?
  - Statistical models of parameters that **vary at more than one level**
  - Multilevel models are particularly appropriate for research designs where data for participants **are organized** at more than one level (i.e., nested data)

- The units of analysis are usually individuals (at a lower level) who are nested within contextual/aggregate units (at a higher level)

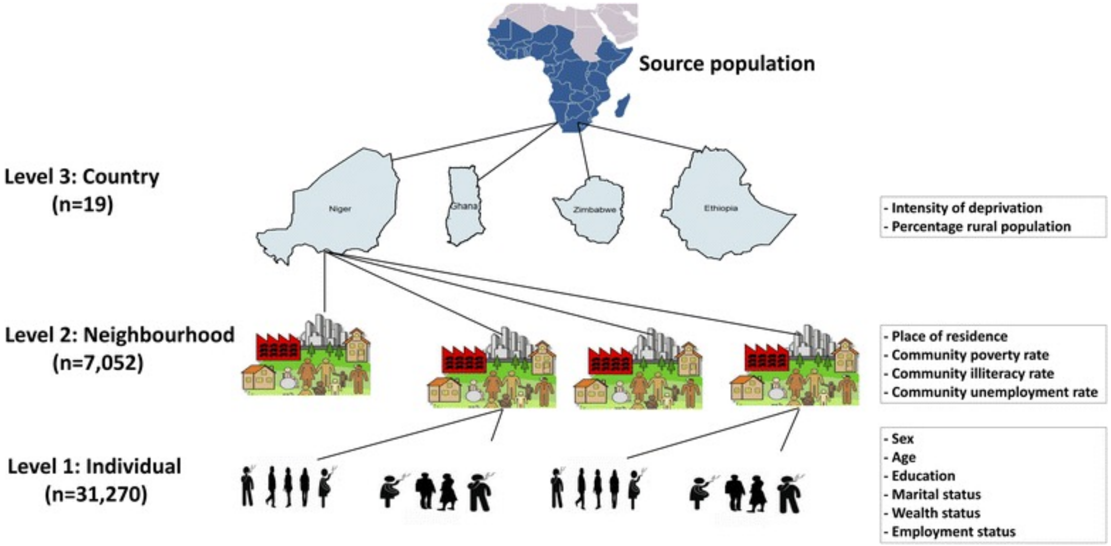


Figure from Uthman, O. A., Ekström, A. M., & Moradi, T. T. (2016). Influence of socioeconomic position and gender on current cigarette smoking among people living with HIV in sub-Saharan Africa: disentangling context from composition. *BMC public health*, *16*(1), 1-9.

Multi-level modeling

- How to perform? (Two stage random effects formulation)
  - First, estimate first-level slope and intercept separately

First-level slope　　　　　　firs-level residual

$$Y_{ij}Employtment\ status = \beta_{0j} + \beta_{1j}Education + \beta_{2j}Martial\ Status + \beta_{3j}Wealth\ status + e_{ij}$$

Intercept

  - Second, first-level beta become the dependent variable of next level analysis

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Place\ of\ residence + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21}Place\ of\ residence + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}Place\ of\ residence + u_{3j}$$

Second-level intercept　　　　Second-level variable

Multi-level analysis

- Benefits of a multilevel approach
  - Correctly account for complex data structures

    : Where single-level model can underestimate each level's variances

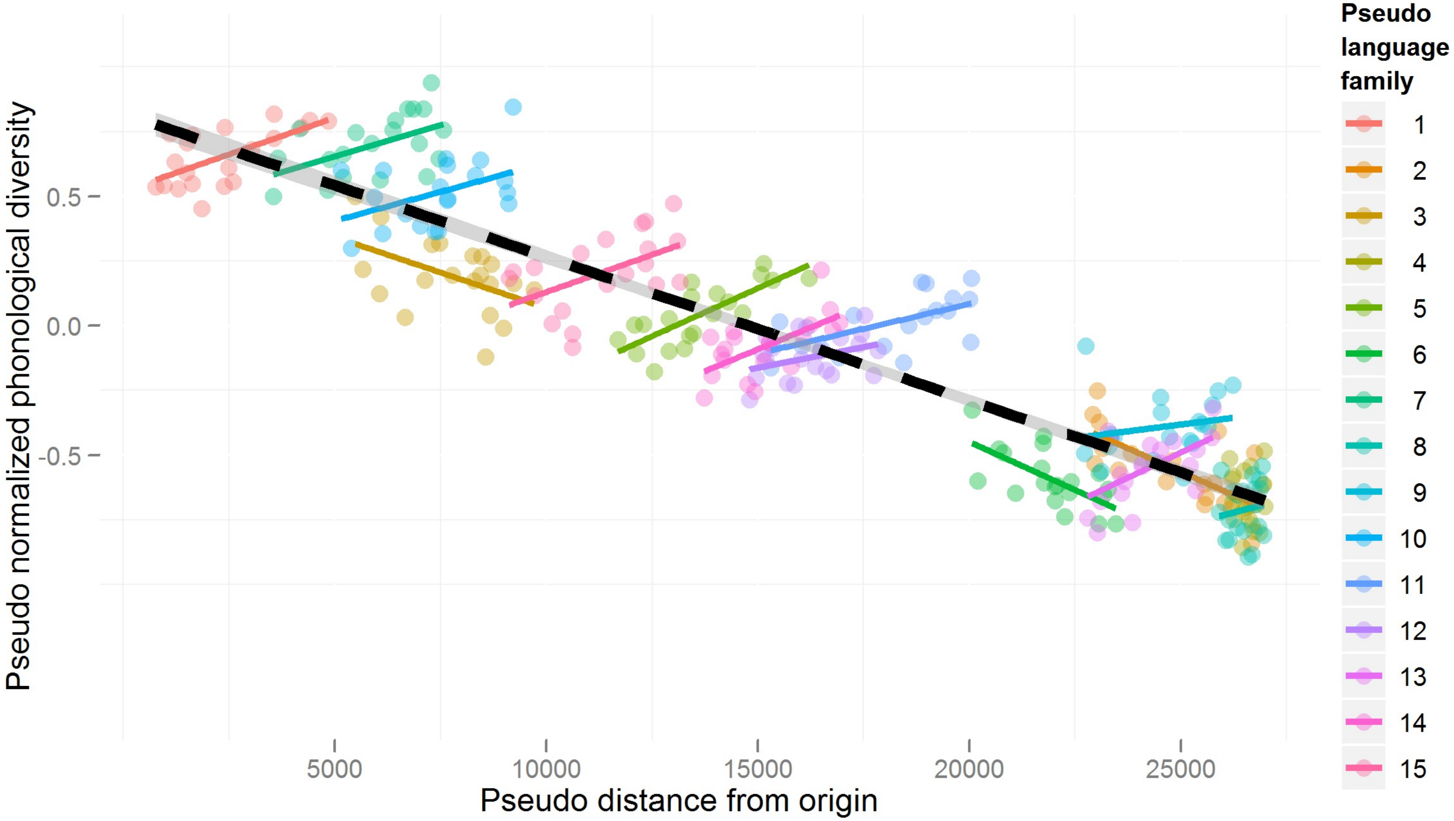
  - Incorporate information on group level relationships

    : Where aggregate analysis only examines group level, and individual analysis can ignore groups (or incorrectly treat group effects as individual effects)

  - Link context to the individual

    : How individual relationships are moderated by broader context

Why do we need multi-level analysis?



Multi-level analysis (visualization-1)

http://mfviz.com/hierarchical-models/?imm_mid=0f7c1e&cmp=em-data-na-na-newsltr_20171108

A Linear Approach

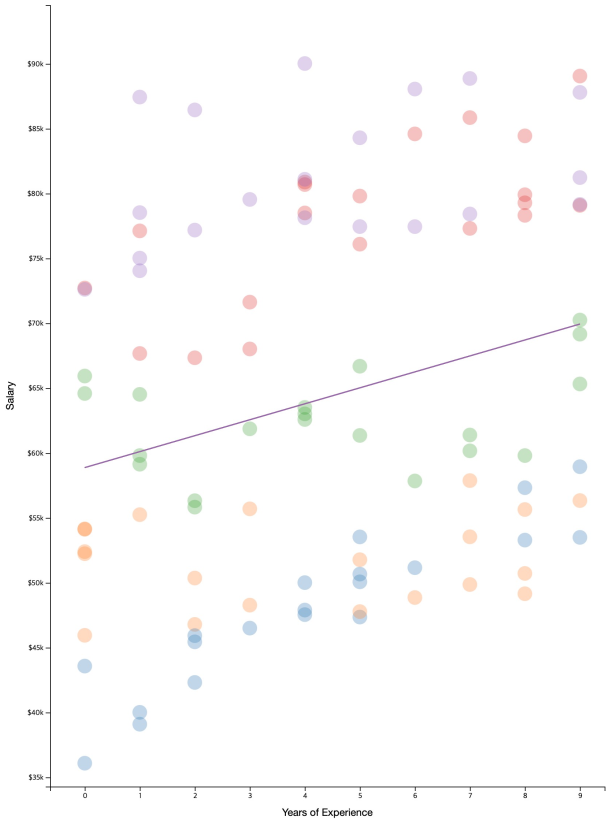
Let's imagine that you're trying to estimate faculty salaries based on their number of years of experience. A simple linear model could be used to estimate this relationship:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

In the above equation, you would estimate the parameters (beta values) for your variables of interest. These are known as the **fixed effects** because they are constant (*fixed*) for each individual. In our case, we would simply use years of experience to predict salary:

$$\hat{salary_i} = \beta_0 + \beta_1 * experience_i$$

While this provides some information about the observed relationship, it is clear that there is variation in salary **by department**. The methods introduced below allow us to capture that variation in different ways.



Random Intercepts

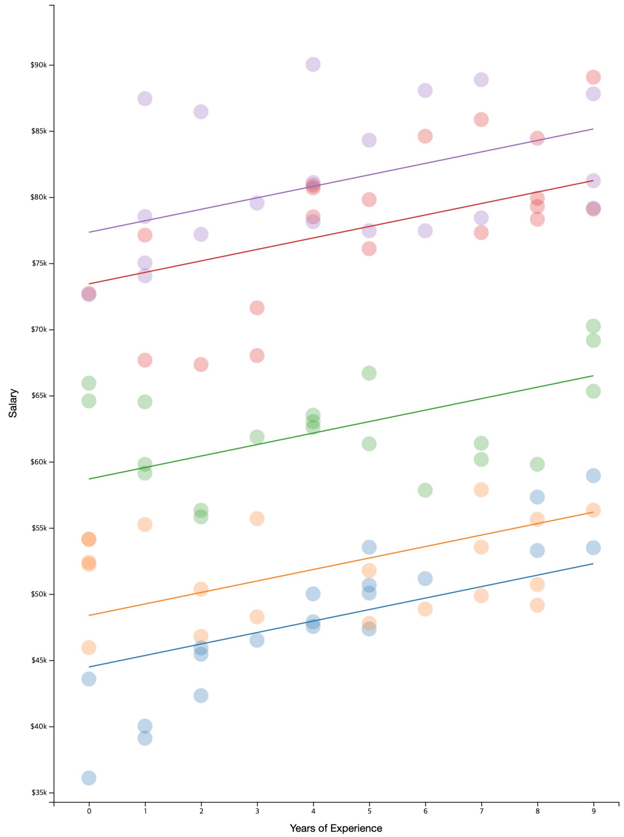
It may be the case that each **department** has a different starting salary for their faculty members, while the annual rate at which salaries increase is consistent across the university. If we believe this to be the case, we would want to allow the **intercept to vary** *by group*. We could describe a **mixed effects** model that allows intercepts to vary by group:

$$\hat{y}_i = \alpha_{j[i]} + \beta x_i$$

In the above equation, the vector of **fixed effects** (constant slopes) is represented by β, while the set of **random intercepts** is captured by α. So, individual `i` in department `j` would have the following salary:

$$\hat{salary_i} = \beta_{0j[i]} + \beta_1 * experience_i$$

This strategy allows us to capture variation in the starting salary of our faculty. However, there may be additional information we want to incorporate into our model.



Multi-level analysis (visualization-2)

http://mfviz.com/hierarchical-models/?imm_mid=0f7c1e&cmp=em-data-na-na-newsltr_20171108

Random Slopes

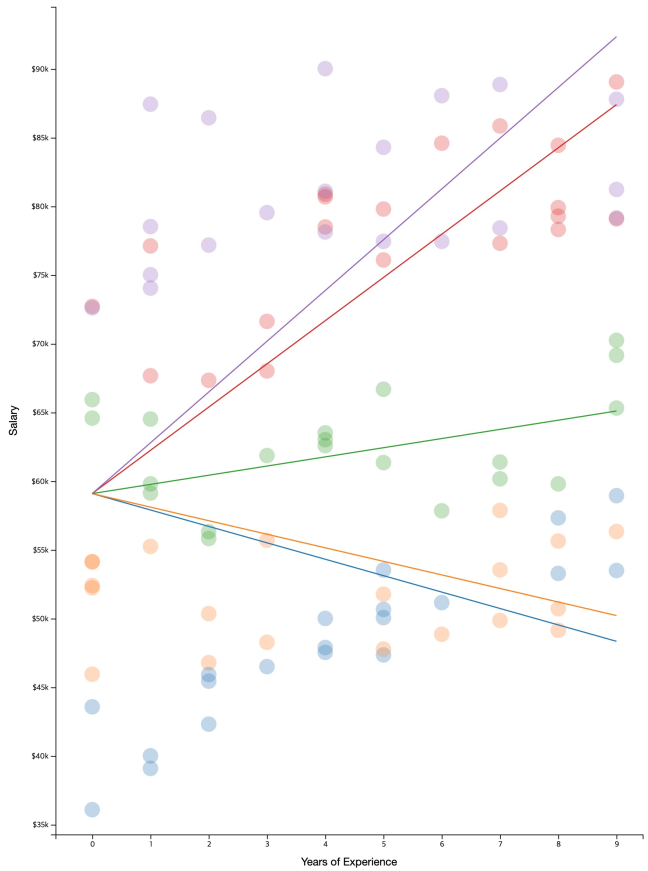
Alternatively, we could imagine that faculty salaries increase at **different rates** depending on the department. We could incorporate this idea into a statistical model by allowing the **slope** to vary, rather than the intercept. We could formalize this with the following notation:

$$\hat{y}_i = \beta_0 + \beta_{j[i]} x_i$$

Here, the intercept (β_0) is constant(fixed) for all individuals, but the slope (β_j) varies depending on the department (`j`) of an individual (`i.`). So, individual `i` in department `j` would have the following salary:

$$\hat{salary}_i = \beta_0 + \beta_{1j[i]} * experience_i$$

While this strategy allows us to capture variation in the *change in salary*, it is clearly a poor fit for the data. We can, however, describe group-level variation in both slope and intercept for a better fitting model.



Random Slopes + Intercepts

It's reasonable to imagine that the most realistic situation is a combination of the scenarios described above:

> Faculty salaries start at different levels *and* increase at different rates depending on their department.

To incorporate both of these realities into our model, we want both the slope and the intercept to vary depending on the department of the faculty member. We can describe this with the following notation:

$$\hat{y}_i = \alpha_{j[i]} + \beta_{j[i]} x_i$$

Thus, the *starting salary* for faculty member `i` depends on their department ($\alpha_{j[i]}$), and their annual raise also varies by department ($\beta_{j[i]}$.):

$$\hat{salary}_i = \beta_{0j[i]} + \beta_{1j[i]} * experience_i$$

In order to implement any of these methods, you'll need to have a strong understanding of the phenomenon you're modeling, and how that is captured in the data. And, of course, you'll need to assess the performance of your models (not described here).