

# Multiple Levels Of The Criminal Mind: Modeling, Profiling & Predicting Serial Killers

Name: Marcus Sinclair

Student ID: 201219978

Module Code: MATH5872M

Module Title: Dissertation in Data Science and Analytics

Supervisor: Dr Benjamin Thorpe

Submission Date: 01/09/2022

## **Abstract**

The curious case of the criminal mind has been researched in great depth by psychologists. Instead of applying a qualitative approach towards understanding serial killers, our research focuses on a quantitative approach, applying the family of statistical models known as multilevel models with the aims to explore whether multilevel modeling is a justified approach to understand the criminal mind. In short, multilevel modeling is a useful tool to use when natural clusters form within the data. A balance of statistical theory coupled with practical applications is seen throughout the research in this dissertation, with data analysed coming from the Radford/FGCU Serial Killer Database- the largest serial killer database in the world. The main findings from this data analysis are the following. Not only is multilevel modeling a justified approach to aid in understanding the criminal mind, but this model framework seems to produce more fruitful results (in specific contexts) in comparison to standard statistical modeling practices such as applying generalized linear models. Multivariate multilevel modeling is subsequently explored with the purpose of determining whether variations seen between serial killers can be used to profile killers, predicting their characteristics and behaviour. Crime-scene data is used with the goal of predicting a killer's race, gender and motive, together with gaining insight into how these characteristics vary within different U.S. states. Our main finding from multivariate analysis is that black serial killers tend to murder victims, on average, within different states to that of white serial killers- variations seen between serial killers do indeed seem to be useful in aiding the profiling of killers.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Objectives . . . . .	6
1.2	Radford/FGCU Serial Killer Database . . . . .	6
<b>2</b>	<b>Generalized Linear Model (GLM)</b>	<b>8</b>
2.1	Application . . . . .	8
2.2	Limitations . . . . .	11
<b>3</b>	<b>Variance Components Model (VCM)</b>	<b>13</b>
3.1	Application . . . . .	13
3.2	Limitations . . . . .	15
<b>4</b>	<b>Random Intercepts Model (RIM)</b>	<b>16</b>
4.1	Application: Imbalanced Covariate . . . . .	16
4.2	Limitations . . . . .	18
4.3	Application: Balanced Covariate . . . . .	19
<b>5</b>	<b>Random Slopes Model (RSM)</b>	<b>21</b>
5.1	Application . . . . .	21
5.2	Limitations . . . . .	23
<b>6</b>	<b>Model Comparison</b>	<b>24</b>
6.1	Likelihood Ratio Test . . . . .	24
6.1.1	Deviance Statistic . . . . .	24
6.1.2	Application . . . . .	25
6.2	Further Model Comparison Metrics . . . . .	26
6.2.1	Akaike Information Criterion (AIC) . . . . .	27
6.2.2	Bayesian Information Criterion (BIC) . . . . .	27
6.2.3	Application . . . . .	27
6.3	Treating Motive As A Fixed Effect . . . . .	28
6.3.1	Limitations . . . . .	29
6.4	Summary . . . . .	30
<b>7</b>	<b>Multivariate Multilevel Models &amp; Profiling Serial Killers</b>	<b>32</b>
<b>8</b>	<b>Binary Responses</b>	<b>34</b>
8.1	Application . . . . .	35
<b>9</b>	<b>Variance Component Multivariate Multilevel Model (VCMM)</b>	<b>38</b>
9.1	Application . . . . .	38
<b>10</b>	<b>Random Intercepts Multivariate Multilevel Model (RIMM)</b>	<b>42</b>
10.1	Application . . . . .	42

<b>11 Conclusion</b>	<b>46</b>
11.1 Multilevel Modeling . . . . .	46
11.2 Multivariate Multilevel Modeling . . . . .	46
11.3 Further Research . . . . .	47
<b>12 Appendix A</b>	<b>49</b>
12.1 (A.1) Statistical Software Used . . . . .	49
12.1.1 Linear Mixed-Effects Models (lmer) Formula Notation . . . . .	49
12.1.2 Multivariate Multilevel Models (runMLwiN) Formula Notation . . . . .	49
12.2 (A.2) RIM and RSM Illustrative Example . . . . .	50
12.3 (A.3) Model Structure Visualisation . . . . .	51
<b>13 Appendix B</b>	<b>52</b>

## List of Figures

1	Age Of First Kill Distribution . . . . .	9
2	Residual Error Assumption Checking . . . . .	10
3	Age At First Kill Clustered By Motive . . . . .	12
4	VCM Quantile-Quantile Plots Of Level 1 & 2 Residual Error . . . . .	13
5	Visualisation Of VCM . . . . .	15
6	RIM (Sex) Quantile-Quantile Plots Of Level 1 & 2 Residual Error . . . . .	17
7	RIM Violin Plot - Unbalanced Covariate . . . . .	18
8	RIM (Race) Quantile-Quantile Plots Of Level 1 & 2 Residual Error . . . . .	19
9	RIM Violin Plot - Balanced Covariate . . . . .	20
10	RSM Quantile-Quantile Plots Of Level 1 & 2 Residual Error . . . . .	22
11	U.S. Clustered By Regions With States Given . . . . .	33
12	Binary RIM Quantile-Quantile Plot Of Level 2 Residual Error . . . . .	35
13	Bar Plot Of Serial Killer Sample Sizes By State . . . . .	36
14	Binary RIM Caterpillar Plot Of Level 2 Residual Error . . . . .	36
15	Binary VCMM Quantile-Quantile Plot Of Level 2 Residual Error . . . . .	39
16	Binary VCMM Caterpillar Plot Of Level 2 Residual Error . . . . .	40
17	Binary RIMM Quantile-Quantile Plot Of Level 2 Residual Error . . . . .	43
18	Binary RIMM Caterpillar Plot Of Level 2 Residual Error . . . . .	43

## List of Tables

1	GLM Estimates For Model (4), AgeFirstKill ~ 1 . . . . .	9
2	GLM Estimates For Model (5), AgeFirstKill ~ 1+Sex . . . . .	11
3	VCM Estimates For Model (6), AgeFirstKill ~ 1 + (1 Motive) . . . . .	14
4	RIM Estimates For Model (9), AgeFirstKill ~ 1+Sex + (1 Motive) . . . . .	17
5	RIM Estimates For Model (10), AgeFirstKill ~ 1+Race + (1 Motive) . . . . .	20

6	RSM Estimates For Model <b>(12)</b> , AgeFirstKill ~ 1+Race + (1+Race Motive) . . . . .	22
7	Likelihood Ratio Test Statistics For GLM, VCM, RIM & RSM . . . . .	25
8	Further Model Comparison Metrics For GLM, VCM, RIM & RSM . . . . .	27
9	GLM Estimates For Model <b>(25)</b> , AgeFirstKill ~ 1+Motive . . . . .	29
10	Data Matrix Example For Multivariate Case . . . . .	32
11	RIM Estimates For Model <b>(31)</b> , logit(Race) ~ 1+Whitevictims + (1 State) . . . . .	37
12	VCMM Estimates For Model <b>(29)</b> , c(logit(Race), logit(Sex), logit(Motive)) ~ 1 + (1 + State) . . . . .	40
13	Data Matrix Of Crime-scene Data . . . . .	42
14	RIMM Estimates For Model <b>(41)</b> , c(logit(Race), logit(Sex), logit(Motive)) ~ 1 + RapedVictims + KilledWithGun + BoundVictims + (1   State) . . . . .	44
15	Variable Outcome Encoding . . . . .	45

# 1 Introduction

## 1.1 Objectives

The main objective of my research is to **demonstrate various multilevel models and how they might be applied in criminology**. We will then subsequently explore variations seen throughout serial killers and show how these findings may be useful to profile serial killers, predicting their behaviour and characteristics. Throughout our discussion, we will build upon our theoretical understanding of multilevel models and the intuition behind such a model framework, with applications to consolidate our understanding seen throughout. The data used has been extracted on the 24th of July 2019 from the Radford/FGCU (Florida Gulf Coast University) database<sup>1</sup>, and is a subset of the worldwide database, with the focus on U.S. serial killers from the 1900s onward. The curious case of the criminal mind has been researched in great depth by psychologists [9] - Instead of applying a qualitative approach towards understanding serial killers, my research focuses on a quantitative approach, with the aims to explore:

1. Is multilevel modeling a justified approach to understand the criminal mind?
2. If so, what variations are prevalent between serial killers with certain traits?
3. Could variations between serial killers analysed be used to profile killers, predicting their behaviour?

## 1.2 Radford/FGCU Serial Killer Database

Before we apply any data analysis, it is important to understand the background of the data that we will be exploring, the reason for its creation, how the data was gathered and how reliable the information is. The Radford/FGCU Serial Killer Database is the largest serial killer database in the world, containing information on 5,334 serial killers worldwide spanning 80 variables.

The idea on constructing this database originated from Dr Michael G. Aamodt, an American industrial and organizational psychology professor at Radford University, Virginia. During an interview with Nick Barksdale<sup>2</sup>, a true-crime journalist and enthusiast, he talked about his time at the University in the early 1980s. He was asked to teach a class on forensic psychology, of-which at the time Dr Aamodt knew fairly little about. He asked his students to produce a serial killer timeline, picking their favourite killer, gathering information on their characteristics and background with the focus on exploring aggressive personality traits. A few years later, he found this past work of his students piled up in one of his cabinets at the University and decided to form a spreadsheet out of the information. Decades later this information was built upon by him and his fellow colleagues at Radford University, eventually resulting in the Radford/FGCU Serial Killer Database, the largest serial killer database in the world.

There are many interesting variables comprising the database for example, data is known regarding the number of victims a serial killer murders in their career, the race, sex and motive; IQ scores and even information on whether a serial killer wet the bed as a child or not. Due to the large span of variables available however, the dataset is extremely porous. As the context may suggest, missing values are prevalent within the database, with not even one serial killer containing information on all the variables. In determining whether the data that is indeed recorded is reliable, four major safeguards were used in its construction [1]:

1. When possible, multiple sources were used for each piece of information.
2. When multiple sources contained conflicting information, more official sources (e.g., state death records, state birth records) were relied on rather than media-driven sources.
3. Graduate students at Florida Gulf Coast University reviewed many of the data fields to corroborate and source the information in each data field.

---

<sup>1</sup>The link to the Radford/FGCU database can be found here: <https://www.fgcu.edu/skdb/>

<sup>2</sup>A video of the interview with Dr Aamodt and Nick Barksdale can be found here: [https://www.youtube.com/watch?v=hfTt1PGp7GU&t=1s&ab\\_channel=CrimeCountry](https://www.youtube.com/watch?v=hfTt1PGp7GU&t=1s&ab_channel=CrimeCountry)

4. As a condition of having access to the database, researchers agreed to provide the founders with new information that was uncovered and to notify them of any data errors they encountered.

Whilst the entire database contains many missing values, data gathered on fundamental serial killer statistics such as the number of victims of the killer, the race, sex and motive is fairly complete with few missing values and accurate. This is especially true for U.S. only serial killers, as the data gathered for serial killers who killed in the U.S. has been cross-referenced extensively by graduate students at Florida Gulf Coast University, with organized state-level records available, dating from the 1900s to the present [1]. During our exploration of the criminal mind, we will only consider U.S. killers who practiced from the 1900s on-wards as this consists of serial killers with the most accurate and reliable information- 1902 serial killers spanning 80 variables.

## 2 Generalized Linear Model (GLM)

We start our discussion on multilevel modeling by first considering the **generalized linear model (GLM)**, a more common and well-known modeling tool in the statistical community. Reasoning for this stems from the GLM being regarded as a specific constrained variant of multilevel modeling [16]. A generalized linear model is any model where the expectation of the output  $Y$  is a function of some linear combination of the inputs  $\beta X = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ , where  $\beta_0, \dots, \beta_p \in \mathbb{R}$  are input parameter constants [5]. That is:

$$\mathbb{E}(Y|X_1 = x_1, \dots, X_p = x_p) = f(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p), \text{ where } f() \text{ is referred to as the link function.} \quad (1)$$

For example, in linear regression,  $f()$  is defined as the identity function- a function that returns the value that was used as its argument. This results in the model:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + e_i, \quad (2)$$

where  $e_i \sim N(0, \sigma^2)$  are referred to as the random error or residuals of the model and are assumed to be **independent and identically distributed (i.i.d)** copies  $(e_i)_{i=1, \dots, n}$ .

### 2.1 Application

In this motivating example we focus on illustrating the limitations of generalized linear models applied to serial killers and how even the most simple multilevel model, the variance components model, introduced in section (3), may be more applicable and have a greater potential to yield fruitful results from the data.

Consider the following questions:

1. How does the age of a serial killer vary when they commit their first murder?
2. Does the distribution of these lengths depend on the gender of the serial killer? For example, is a killer's age at first kill, on average, higher for females rather than male killers?

The age a serial killer murders their first victim is given to the nearest year by **AgeFirstKill**, a numerical variable with known records for 93% of the serial killers recorded in the Radford/FGCU database, 1763 out of the 1902 individuals.

With the focus on our first question regarding variability between **AgeFirstKill** lengths, let us assume the duration periods are independent and identically distributed, with normal distribution. Thus we propose the model:

$$y_i \sim N(\beta_0, \sigma^2), \text{ with i.i.d copies } (y_i)_{i=1, \dots, n}, \quad (3)$$

where  $y_i$  denotes the age at first kill of the  $i$ -th serial killer, for individuals  $i \in 1, \dots, n = 1763$ . Equivalently, we can write this model as

$$y_i = \beta_0 + e_i, \text{ with } e_i \sim N(0, \sigma^2), \text{ i.i.d copies } (e_i)_{i=1, \dots, n}, \quad (4)$$

where each  $e_i$  denotes the residual of the  $i$ -th serial killer. It is useful to contextualize equation (4) as the generalized linear model with normal link function (2) such that  $x_{1i} = \dots = x_{pi} = 0$ , i.e., no inputs are present. The purpose of writing the model in this format will become apparent following generalized linear model extensions to multilevel models in the following section. Before we answer our questions by applying model (4), we must first check the model assumptions are met that is, normality is present in our residuals  $e_i$ . Since the residuals are directly proportional to the output **AgeFirstKill**, looking at this variable seems like the natural approach here.



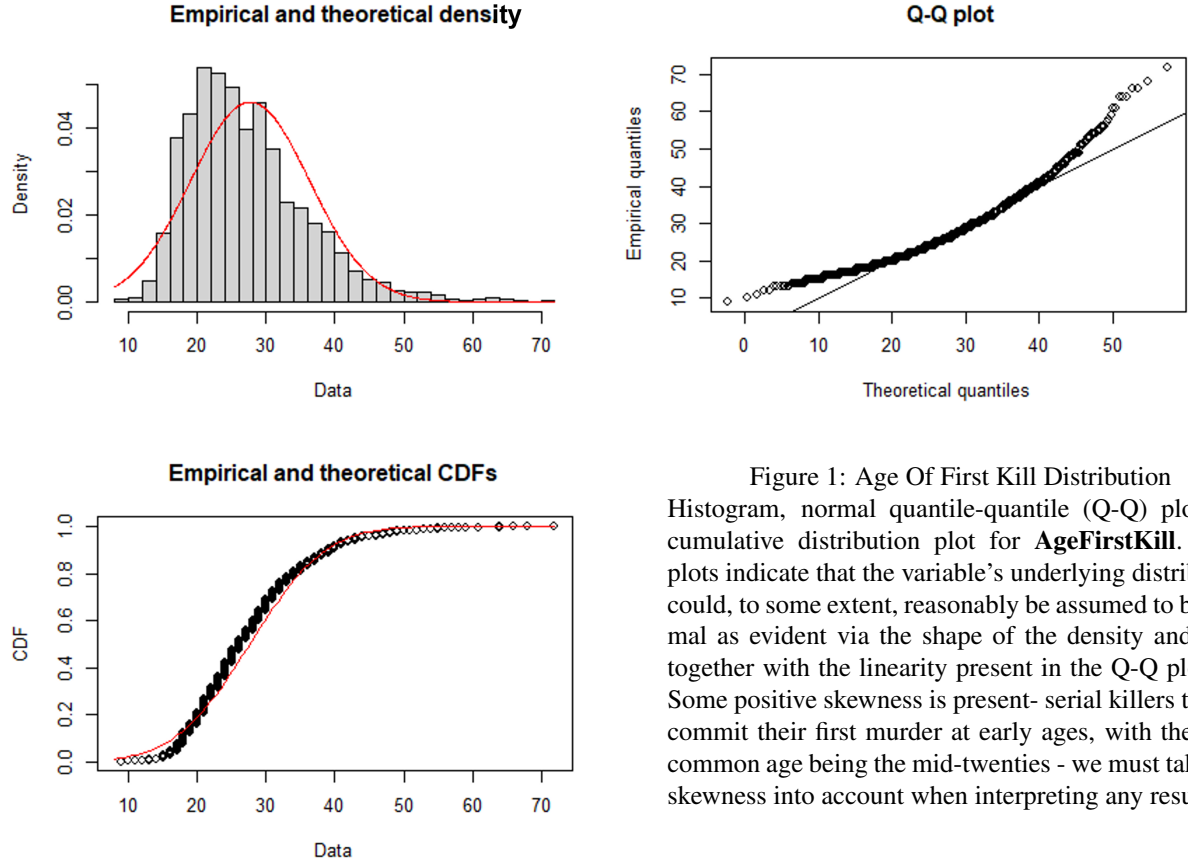


Figure 1: Age Of First Kill Distribution  
Histogram, normal quantile-quantile (Q-Q) plot and cumulative distribution plot for **AgeFirstKill**. The plots indicate that the variable's underlying distribution could, to some extent, reasonably be assumed to be normal as evident via the shape of the density and CDF together with the linearity present in the Q-Q plot [2]. Some positive skewness is present- serial killers tend to commit their first murder at early ages, with the most common age being the mid-twenties - we must take this skewness into account when interpreting any results.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	27.64	27.24-28.05	<0.001
standard error	8.67		
Observations	1763		

Table 1: GLM Estimates For Model (4), AgeFirstKill ~ 1

The model estimates our unknown parameter  $\beta_0$  by the method of ordinary least squares (finding the  $\beta_0$  that minimises  $\sum_{i=1}^n e_i^2$ ). Given individuals  $n = 1763$ , this produces estimates of the population mean,  $\hat{\beta}_0 = 27.64$  and residual standard error  $\hat{\sigma}^2 = 8.67^2$ . From table (1)<sup>a</sup>, we see a 95% confidence interval (CI) and p-value (p) for the intercept parameter  $\beta_0$ . These are defined in the usual sense for instance, the confidence interval is constructed by considering the **estimate**  $\pm$  **standard error**  $\times t_{2.5\%, n-1}$ , where  $t_{2.5\%, n-1}$  is the 97.5% quantile of the t-distribution with  $n - 1$  degrees of freedom [21].

<sup>a</sup>For more information on the meaning behind formula notation seen in the caption of table (1) and the statistical software used throughout model applications, see appendix (A).

Here,  $n = 1763$  denotes the sample size used in model construction. Note, the degrees of freedom in our interval can be written in a more general sense as  $n - p - 1$ , with  $p$  referring to the number of inputs present as referenced prior in equation (2). The p-value measures the probability of obtaining a regression coefficient result at least as extreme as the observed result produced by the GLM- it is typically agreed in the statistical community that a p-value  $< 0.05$  is deemed to be useful indicator showing the input as a predictor of the output [19]. The age a serial killers commits

their first murder is, on average, seen to occur in their mid-twenties with two-thirds of all killers starting their killings between twenty and forty years of age, with the remaining tailing off at the younger and older years.

Regarding the second question on the possible dependence of the age at first kill and the gender of a serial killer, we would usually estimate a linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \text{ with } e_i \sim N(0, \sigma^2), \text{ i.i.d copies } (e_i)_{i=1, \dots, n}, \quad (5)$$

where  $x_i$  denotes the gender of the  $i$ -th serial killer. This input is given by the variable **Sex**, a categorical variable that takes binary values (*Male*, *Female*) with (1745, 150) individuals respectively. Only 7 individuals in the Radford/FGCU database have an unknown gender. It may seem strange to apply a categorical variable to a linear regression problem (taking residuals of discrete values seems somewhat nonsensical at first impression) however, given the binary nature of the output of **Sex** and the fact that **Sex** is the only feature input present in the regression model, application of such a model will lead to interpretable and useful results [20]. The variable is encoded with numerical values mapped as (*Male*, *Female*)  $\rightarrow$  (0, 1), with parameter estimates calculated via minimizing the ordinary least squares of the residuals as usual. Again, standard practice dictates to check the model assumptions of (5) prior to analysis that is, normality is present within the residuals  $e_i$ .

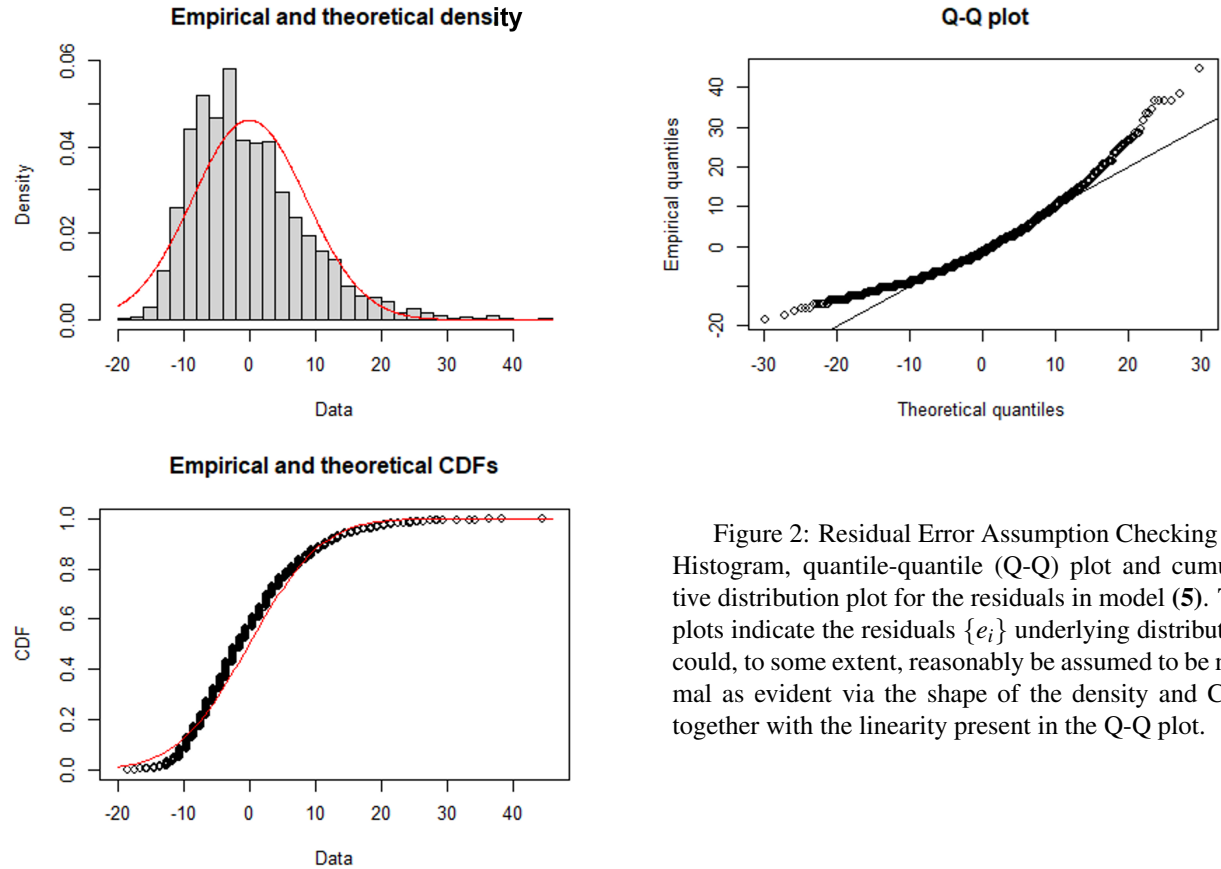


Figure 2: Residual Error Assumption Checking Histogram, quantile-quantile (Q-Q) plot and cumulative distribution plot for the residuals in model (5). The plots indicate the residuals  $\{e_i\}$  underlying distribution could, to some extent, reasonably be assumed to be normal as evident via the shape of the density and CDF together with the linearity present in the Q-Q plot.

In regards to the distribution of the residuals  $\{e_i\}$ , it is worth noting that, similar to the distribution of **AgeFirstKill**, a substantial amount of positive skewness is present within the residuals as evident via the positive curvature seen at both extremes of the theoretical quantiles in the Q-Q plot and shape of the histogram in figure (2) - our model results should again be interpreted with care.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	27.48	27.05-27.90	<0.001
Sex[Female]	2.12	0.62-3.62	0.006
Standard error	8.65		
Observations	1763		
R <sup>2</sup>	0.004		
F-statistic	7.71		

Table 2: GLM Estimates For Model (5), AgeFirstKill ~ 1+Sex

Again, estimating via ordinary least squares results in parameter estimates  $\hat{\beta}_0 = 27.48$  and  $\hat{\beta}_1 = 2.12$ , with an estimate of the residual standard error,  $\hat{\sigma}^2 = 8.65^2$ . Given the presence of a feature input **Sex** in the model, we now have some new statistical objects seen in table (2) to consider that is, the R-squared and F-statistic values. An R-squared value helps determine how well the regression models explains the observed data [20]. For instance, here we find only 0.4% of the variability in the age of first kill of a serial killer is explained by the gender of the killer. Female killers, on average, tend to start two years later than their male counterparts.

The large F-statistic of 7.71, coupled with the small p-value  $p = 0.006 \ll 0.05$  and confidence interval  $CI = [0.62, 3.62]$  situating in the positive domain, suggests that this result is significant throughout serial killers. The distribution of the lengths of age at first kill does indeed depend on the gender of the killer albeit a small amount. A killer’s age at first kill is, on average, higher for females rather than male killers.

## 2.2 Limitations

Application of this generalized linear model approach seems reasonable thus far however, there are at least three issues regarding this methodology:

- We have **assumed independence**. Our sample of  $n = 1763$  serial killers may be correlated in relation to some feature variable present in the data. For instance, do you think that the average age at first kill of a serial killer remains constant throughout say, different serial killer motives? It seems natural to hypothesize motives such as *Financial/personal gain* or *Black widow* (the killing of spouses) to occur at a later age in comparison to motives such as *Organised crime* or *Cult-related* serial killings. If such correlation between individuals is present then a GLM assumption is violated resulting in unreliable findings.
- We probably haven’t answered the questions as they were intended. For example, the wording “vary between serial killers” is not very clear. Do we want to know about the killer-to-killer variability in age at first kill? Or do we want to know about the variability in age at first kill with respect to some feature input such as **Race**, **Motive** or **KilledWithAccomplice**? Or do we want to know about both?
- Are our normality assumptions justified? We have applied two generalized linear models, both of which assumed normality within the residuals. As seen by our assumption investigations in figures (1) and (2), these normalities assumed could be questioned. Both distributions, that of **AgeFirstKill** and the residuals  $\{e_i\}$  in model (5) have a substantial amount of positive skewness. Real-world data is never “perfect” and some liberties must be taken in regards to fitting to model assumptions. It is up to the statistician to determine if the amount of such liberties taken is justified given the context of the hypothesis in question [6].

Let's investigate the independence assumption for our first model (4) for which we estimated the average age at first kill to be  $\hat{\beta}_0 = 27.64$  years.

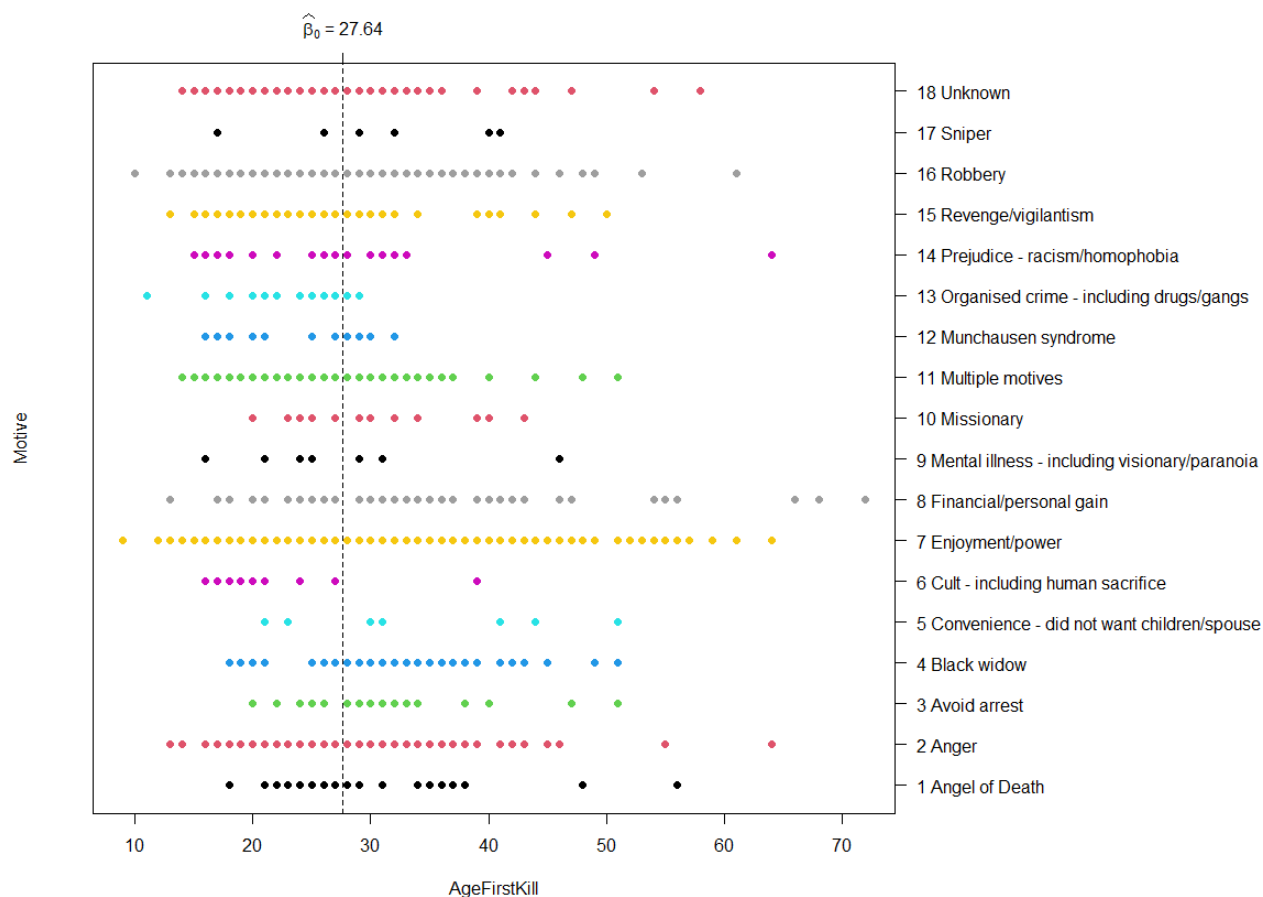


Figure 3: Age At First Kill Clustered By Motive

Plot of **AgeFirstKill** clustered by **Motive**, a categorical variable taking 18 unique values as seen in the plot, from *Angel of Death* to *Unknown* motives. The estimated age at first kill across all killers is given by the dashed vertical line, with a value of  $\hat{\beta}_0 = 27.64$  years. It appears that the average duration varies between killers with different motives.

As seen in figure (3), the age at first kill may be correlated in relation to the motive of the killer. As hypothesised prior, serial killers relating to *Organised crime* seem to start killings at an earlier age than average. Vice versa, *Financial/personal gain* incentives and the *Black widow* killings occur at later ages than average. These natural clusterings seen are ignored in a standard generalized linear model framework. As such, our independence assumptions of the response variable  $\{y_i\}$  made in the generalized linear model case could be argued to be invalidated [4].

We now introduce the main focus of our discussion that is, multilevel models- a tool that can extend the generalized linear model framework to internalise such groupings present in the data.

### 3 Variance Components Model (VCM)

Multilevel modeling is a strategy that internalises the clustered nature of our data within the model. The idea is to produce a framework with multiple levels where level 1 refers to individuals or observations as seen in the standard generalized linear model, and level 2 refers to the cluster-level whereby certain individuals behave similarly within respective clusters [4]. This framework could be built upon in perpetuity for instance, a level 3 super-cluster could be internalised and so-forth. For simplicity, let us first consider the level 2 model with the absence of inputs/covariates  $X$ :

$$y_{ij} = \beta_0 + u_{0j} + e_{0ij}, \quad (6)$$

with  $u_{0j} \sim N(0, \sigma_{u0}^2)$  i.i.d copies  $(u_{0j})_{j=1, \dots, m}$  and  $e_{0ij} \sim N(0, \sigma_{e0}^2)$  i.i.d copies  $(e_{0ij})_{i=1, \dots, n, j=1, \dots, m}$  (see previous application for the meaning of i.i.d). Here,  $u_{0j}$  is a level 2 (cluster-level) error formally refereed to as the random effect of the  $j$ -th group on output/response variable  $y_{ij}$ . The value of  $u_{0j}$  is assumed to be the same for all individuals pertaining to the  $j$ -th cluster, and is assumed to be independent to level 1 random errors  $e_{0ij}$ . The zero subscript used in level 1 and 2 errors  $e_{0ij}$  and  $u_{0j}$  may seem somewhat cumbersome however, its inclusion allows model extension to appear intuitive and complete. A model of the form given in equation (6) is refereed to as a **variance components model (VCM)**.

#### 3.1 Application

Recall the first question given in section (2.1) :

1. How does the age of a serial killer vary when they commit their first murder?

Given this question, our goal here is to determine whether VCM may be more applicable and have a greater potential to yield fruitful results from the data compared to the GLM application. Let  $y_{ij}$  denote the age at first kill corresponding to the  $i$ -th serial killer (the  $i$ -th level 1 individual) of the  $j$ -th cluster (level 2). Define level 2 clusters as the motives seen and labeled in figure (3) i.e.,  $j = 1$  refers to the motive *Angel of Death*,  $j = 2$ , *Anger* and so forth up to  $j = 18$ , *Unknown* motive. Before applying the VCM we must first check our model assumptions that is, normality of not just the level 1 standard error residuals  $\{e_{0ij}\}$ , but we must find normality present within the level 2 random error  $\{u_{0j}\}$  too.

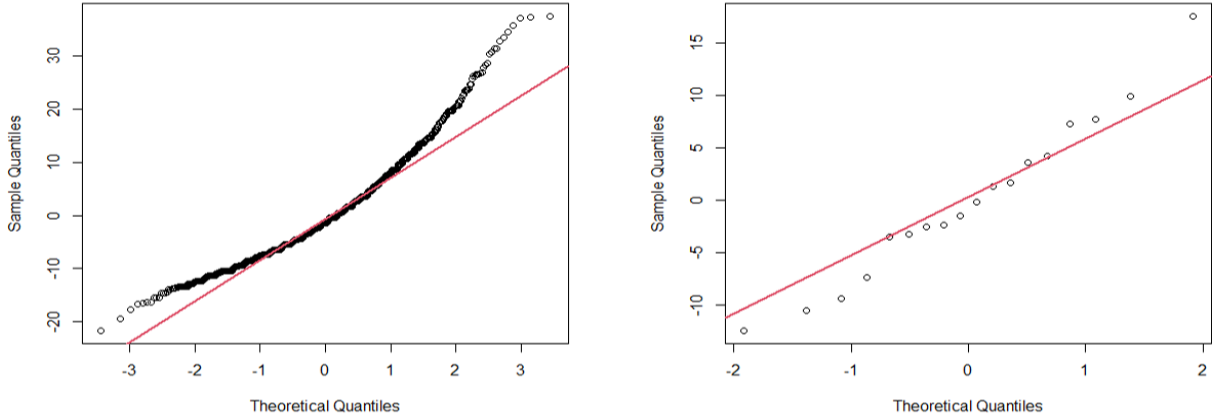


Figure 4: VCM Quantile-Quantile Plots Of Level 1 & 2 Residual Error

Left: Level 1 standard error residuals  $\{e_{0ij}\}$  quantile-quantile plot. Similar issues (as expected) arise for assumption checking for model (6) to that seen in application (2.1), that of positive skewness present at the level 1 state. Right: Level 2 random error  $\{u_{0j}\}$  quantile-quantile plot. 18 points are seen which are reference to the 18 unique values the variable **Motive** can take. Linearity at the level 2 state is seen throughout the points - normality model assumption is justified.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	27.87	26.19-29.56	<0.001
<b>Random Effects</b>			
$\sigma_{e0}^2$	8.41 <sup>2</sup>		
$\sigma_{u0Motive}^2$	3.15 <sup>2</sup>		
ICC	0.13		
$N_{Motive}$	18		
Observations	1763		

Table 3: VCM Estimates For Model (6), AgeFirstKill ~ 1 + (1|Motive)

As shown in table (3), applying VCM yields the estimate  $\hat{\beta}_0 = 27.87$  years for the population mean  $E(y_{ij}) = \beta_0$ . Unlike in the generalized linear model case whereby one variance parameter is estimated, for the variance components model we require to estimate two:  $\hat{\sigma}_{u0}^2 = 3.15^2$  at the motive cluster level, and  $\hat{\sigma}_{e0}^2 = 8.41^2$  at the killer individual level. These suggest that while most variability in the age at first kill of a serial killer stems from differences between killers, a substantial amount of variability is explained by the motive behind such killings. An interesting and potentially surprising finding is that the population mean for VCM marginally differs from that of the generalized linear model applied in section (2),  $\hat{\beta}_0 = 27.87$  compared to that of 27.64 years seen prior.

Whilst the generalized linear model applied in section (2) estimates model parameters via maximum likelihood methods i.e, for linear regression, by the ordinary least squares method (OLS); the variance components model uses a more general framework to estimate model parameters- the iterative generalized least squares (IGLS) method. The details regarding IGLS will not be discussed here, but can be found in Goldstein’s book on multilevel modeling [6].

Parameters estimated via the iterative generalized least squares method will converge to equivalent parameter values produced via maximum likelihood. The differences in parameter values  $\hat{\beta}_{0VCM} = 27.87$  compared to that of  $\hat{\beta}_{0GLM} = 27.64$  is due to the difference in parameter estimation methods.

Additional statistics will be examined during our multilevel modeling journey with the first being the intraclass correlation coefficient (ICC) seen in table (3) taking a value of 0.13 for the variance components model applied. This statistic is a measure of the correlation between two different individuals in the same cluster,  $corr(y_{ij}, y_{kj})$  for  $i \neq k$  [4]. It describes how strongly individuals in the same cluster resemble each other, taking a value from 0 to 1 with larger values indicating a high resemblance in the cluster. For a variance components model, the intraclass correlation coefficient can be calculated by

$$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2} = \frac{3.15^2}{3.15^2 + 8.41^2} = 0.13, \text{ for the example above.} \quad (7)$$

The formula makes intuitive sense for VCM as the intraclass correlation coefficient is just the proportion of level 2 variation seen in regards to the total variance at levels 1 and 2. Whilst interpreting parameter estimates and statistics is important and useful for deepening our understanding of serial killers, we should be careful about blindly applying models, inferring their results, but focus on the mechanism behind model construction too, deepening our understanding of model practices as the following figure illustrates.

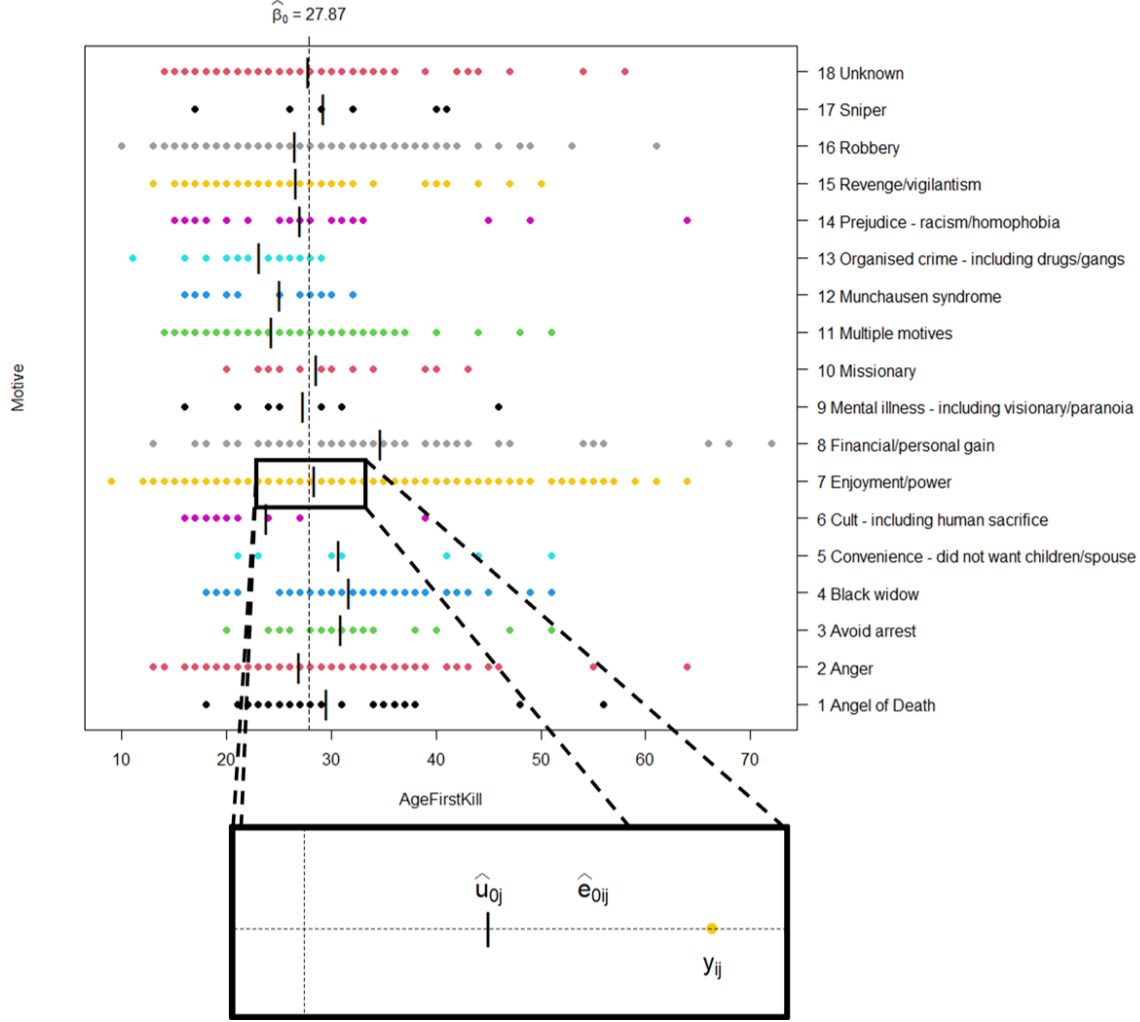


Figure 5: Visualisation Of VCM

Expanded plot to that seen in figure (3) with the parameter estimate  $\hat{\beta}_0 = 27.87$  produced by VCM seen by the dotted line. Black bars are shown for each of the 18 individual motives that represent the estimated cluster means  $\hat{\beta}_0 + \hat{u}_{0j}$ . A grid zoom is shown for an individual  $y_{ij}$  for the motive *Enjoyment/power* that is, for  $j = 7$ . The grid zoom shows the mechanism of model (6). Individuals in a specific cluster are normally distributed with noise  $e_{0ij} \sim N(0, \sigma_{e0}^2)$  around their specific cluster mean - the black bar, with the cluster means being themselves normally distributed with noise  $u_{0j} \sim N(0, \sigma_{u0}^2)$  around the parameter  $\beta_0$ .

### 3.2 Limitations

Hopefully this example has given insight into the usefulness of multilevel modeling. Whilst only the most simple multilevel model (the VCM) has been introduced thus far, more complex extensions are still relatively straight forward and even routine to implement using standard statistical software. Therefore, it is of great importance to reflect on whether model assumptions are met and hierarchical structures are prevalent before blindly applying such models. It is common for standard single level models such as the GLM to suffice. The VCM cannot be applied to hypotheses regarding variability between feature inputs and a response variable, considering an alternative multilevel model that could be applied in such cases will now be explored.

## 4 Random Intercepts Model (RIM)

Here we introduce an extension to the level 2 VCM seen in section (3). Unlike the variance components model, the **random intercept model (RIM)** allows the inclusion of inputs/covariates  $\beta X = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$  where  $\beta_0, \dots, \beta_p \in \mathbb{R}$  are input parameter constants[6]. A random intercept model applied to our sample  $\{y_{ij}\}$  is defined such that

$$y_{ij} = (\beta_0 + u_{0j}) + \sum_{l=1}^p \beta_l x_{lij} + e_{0ij}, \quad (8)$$

where all random terms on the right hand side are mutually independent and normally distributed

$$e_{0ij} \sim N(0, \sigma_{e0}^2) \text{ and } u_{0j} \sim N(0, \sigma_{u0}^2).$$

It is important to understand and distinguish the fixed and random effects in multilevel models.  $\beta_l$  is the constant fixed effect of the  $l$ -th covariate  $l \in 1, \dots, p$ .  $u_{0j}$  is the level 2 random effect of cluster  $j$ .  $e_{0ij}$  is the level 1 random error for the  $i$ -th individual in cluster  $j$ . Since the level 2 random effect  $u_{0j}$  is constant within a specific cluster  $j$ , we can visualise a random intercepts model as  $m$  regression models<sup>3</sup>, one for each cluster, with different intercepts  $\beta_0 + u_{0j}$ . We have  $p + 3$  unknown, constant parameters to estimate: fixed effects  $\beta_0, \dots, \beta_p$ , and variance parameters  $\sigma_{e0}^2$  and  $\sigma_{u0}^2$ .

### 4.1 Application: Imbalanced Covariate

Recall the second question given in section (2.1):

1. Does the distribution of the age at first kill depend on the gender of the serial killer?

Our goal here is to show the usefulness of a random intercepts model and how model results vary in relation to its level 1, generalized linear model counterpart. Applying RIM is a good choice here. Justification for applying the random intercepts model stems from clustering in our data. Normally, we would fit just the single level regression model as applied in section (2.1). This is ill-advised because of the grouping present within the data. Furthermore, the question above is about means, and we can answer it using the fixed part, that is, the slope  $\beta_1$  of the overall regression line. This is analogous to fitting a standard single level regression model where we'd use our estimate of  $\beta_1$  to answer the question.

It seems reasonable to suggest a variance components model could be used regarding the question given above however, VCM is not applicable here because we need to control for the differences in the feature input **Sex**. RIM does indeed allow control for those differences. We can answer the question using the random part, the level 2 variance,  $\sigma_{u0}^2$ .

We propose the model:

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + e_{0ij}, \quad (9)$$

where  $y_{ij}$  denotes the **AgeFirstKill** of the  $i$ -th serial killer in the  $j$ -th cluster, **Motive**. This model is equivalent to (8) with only one covariate present  $x_{1ij}$ , the **Sex** of the  $i$ -th serial killer in the  $j$ -th cluster.

---

<sup>3</sup>see appendix (A.2) for illustrative example.



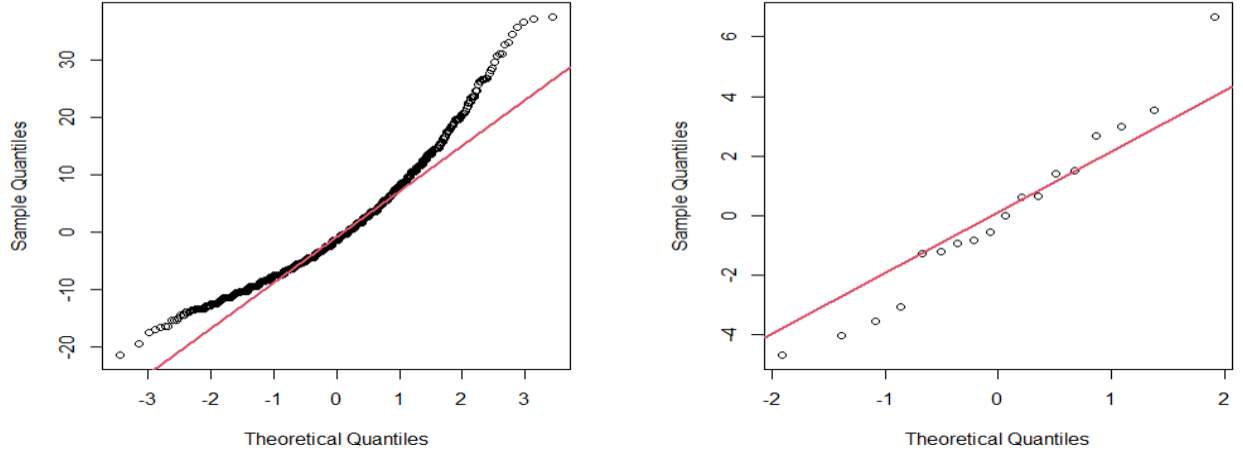


Figure 6: RIM (Sex) Quantile-Quantile Plots Of Level 1 & 2 Residual Error

Left: Level 1 standard error residuals  $\{e_{0ij}\}$  quantile-quantile plot. Right: Level 2 random error  $\{u_{0j}\}$  quantile-quantile plot. 18 points are seen which are reference to the 18 unique values the variable **Motive** can take. Only a marginal difference is seen in the distribution of RIM residuals in comparison to VCM residuals - I refer you to figure (4) for implications.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	27.76	26.05–29.46	<0.001
Sex[Female]	0.52	-1.21–2.24	0.558
<b>Random Effects</b>			
$\sigma^2_{e0}$	8.41 <sup>2</sup>		
$\sigma^2_{u0Motive}$	3.19 <sup>2</sup>		
ICC	0.13		
$N_{Motive}$	18		
Observations	1763		

From figure (6), it seems justifiable to assume the underlying distribution of this variable to be normal, thus the assumption of normality of our sample  $\{y_{ij}\}$  required by RIM is met. Application of the random intercepts model yields parameter estimates  $\hat{\beta}_0 = 27.76$  and  $\hat{\beta}_1 = 0.52$  however, unlike in our GLM example in section (2.1) whereby we observed a meaningful impact of the gender of a killer on the age at first kill, the analogous RIM version does not. Our confidence interval  $CI = [-1.21, 2.24]$  spans a negative and positive domain. Our p-value  $p = 0.558$  is large.

Table 4: RIM Estimates For Model (9), AgeFirstKill  
~ 1+Sex + (1|Motive)

Given the similarities in both models, it may seem strange to see such a large disparity in terms of the significance of our findings. What limitation of multilevel modeling have we just stumbled across in this example?

## 4.2 Limitations

Plot showing model (9) for the most prevalent motives *Enjoyment/power* and *Robbery*. Violin plots are seen, two for each motive with the raw data points shown by the black dots. For a given age of first kill, the greater the width of the violin plot the greater the density of serial killers - (750,9) individuals (*Male, Female*) for *Enjoyment/power* and (367,18) individuals (*Male, Female*) for *Robbery*.

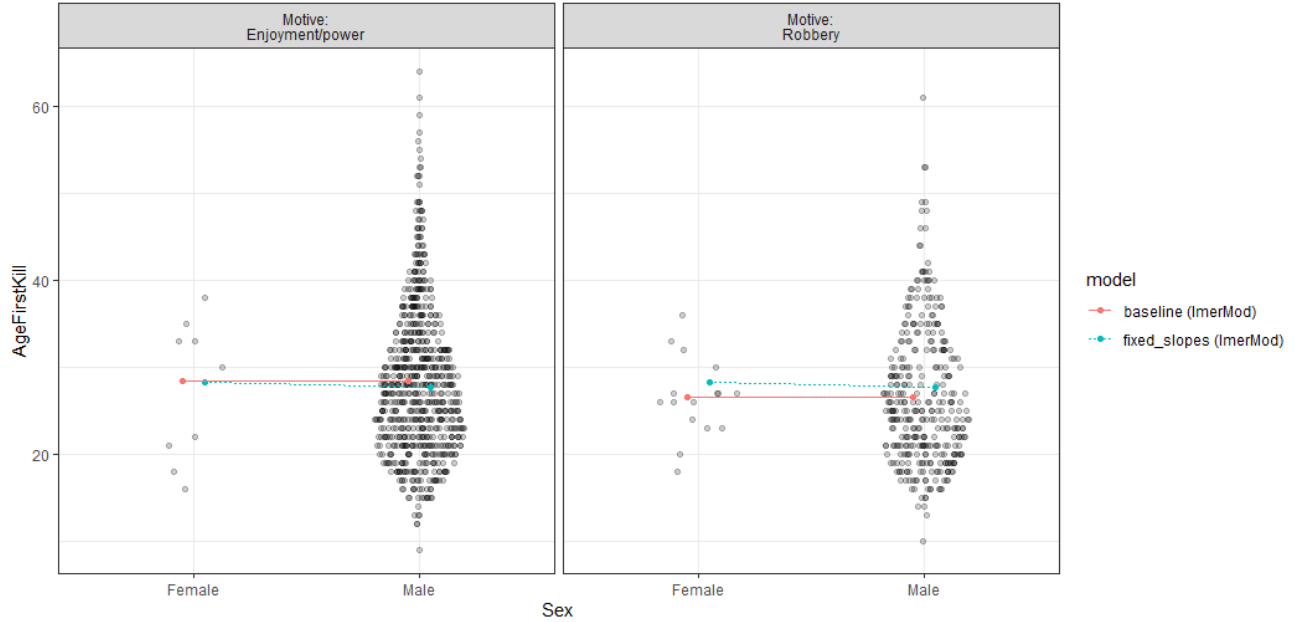


Figure 7: RIM Violin Plot - Unbalanced Covariate

The red line shows a VCM for **AgeFirstKill** and is referred to as a baseline model. The blue line is the RIM in question as is referred to as a fixed\_slopes model (see appendix (A.1) for meaning behind lmerMod in figure (7)). The violin plots in the figure are an example of the large imbalance within the value of our covariate **Sex**. Throughout the entire data set we find a large dominant presence in male serial killer in comparison to female with (*Male, Female*) comprising of (1745,150) individuals respectively.

RIM derives its parameters in a similar fashion to that of the VCM of which, we have seen the mechanism behind its construction- see figure (5). Large data imbalances results in large uncertainties in estimations of our level 1 and 2 residuals  $\{e_{0ij}\}$  and  $\{u_{0ij}\}$ . For instance in relation to the RIM model seen in figure (7), consider a new addition to the data for example, a female serial killer whom started their killings at the old age of 60.

This finding would impact our parameter estimates of  $\beta_0$  and  $\beta_1$  to a greater extent in comparison to a new male serial killer observation of the equivalent age. Imbalances in data, especially with small samples within clusters present as seen in this example brings uncertainty in our findings. For such an imbalanced feature input, partitioning the data into various clusters risks small sample sizes within specific groups. Whilst this limitation is applicable to regression models in general, the groupings produced from a multilevel modeling approach emphasizes this limitation to a greater extent to that which is seen in standard regression. To consolidate this point, let us now apply a more balanced feature input of interest for instance, **Race**, as opposed to **Sex** investigating whether this issue still arises.

### 4.3 Application: Balanced Covariate

Consider the following hypothesis:

1. The distribution of the age at first kill varies within the race of a serial killer. Black ethnic groups tend to start killings at an earlier age compared to white serial killers.

Our goal in this example is to apply equivalent analysis to that seen in section (4.1) however, analysis is now applied upon a more balanced feature input **Race** - A categorical variable with values (*White, Black, Hispanic, Asian, Native American*) with killer samples (1024, 750, 90, 10, 13) respectively. Analogous to the variable **Sex** taking binary values (*Male, Female*), we transform **Race**, removing the minority classes *Hispanic, Asian* and *Native American*, treating them as missing values.

Justification stems from the small sample sizes present in the data for such classes, coupled with our goal of reducing down the feature input **Race** to that of a binary variable, similar to the variable **Sex** as analysed prior. From an imbalanced covariate **Sex** with values (*Male, Female*) with (1745,150) individuals, we are ready to consider the more balanced covariate **Race** with values (*White, Black*) with (1024, 750) individuals. We propose the model:

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + e_{0ij}, \quad (10)$$

where  $y_{ij}$  denotes the **AgeFirstKill** of the  $i$ -th serial killer in the  $j$ -th cluster, **Motive**. This model is equivalent to (8) with only one covariate present  $x_{1ij}$ , the **Race** of the  $i$ -th serial killer in the  $j$ -th cluster. Let us check our model assumptions.

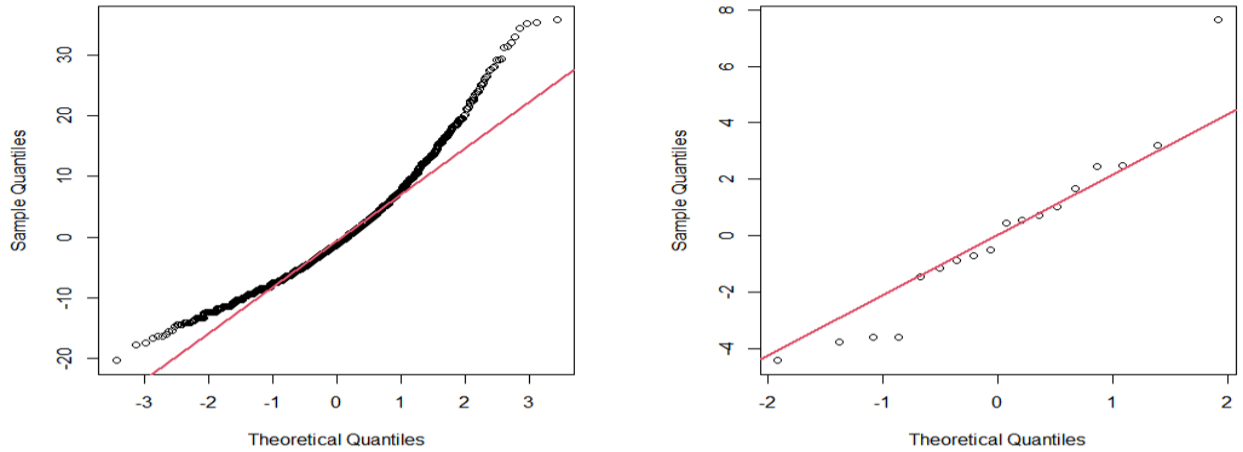


Figure 8: RIM (Race) Quantile-Quantile Plots Of Level 1 & 2 Residual Error

Left: Level 1 standard error residuals  $\{e_{0ij}\}$  normal quantile-quantile plot. Right: Level 2 random error  $\{u_{0j}\}$  normal quantile-quantile plot. Only a marginal difference is seen in the distribution of RIM residuals in comparison to VCM residuals - I refer you to figure (4) for implications.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	26.39	24.60–28.18	<0.001
Race[White]	2.63	1.77–3.49	<0.001
<b>Random Effects</b>			
$\sigma^2_{e0}$	8.38 <sup>2</sup>		
$\sigma^2_{u0Motive}$	3.27 <sup>2</sup>		
ICC	0.13		
$N_{Motive}$	18		
Observations	1649		

Table 5: RIM Estimates For Model (10), Age-FirstKill  $\sim 1 + \text{Race} + (1|Motive)$

Black ethnic groups do indeed tend to start killings at an earlier age compared to white serial killers, with the data suggesting around two and a half years difference between the two groups. Our confidence interval  $CI = [1.77, 3.49]$  spans a positive domain. Our p-value  $p < 0.001$  is sufficiently small.

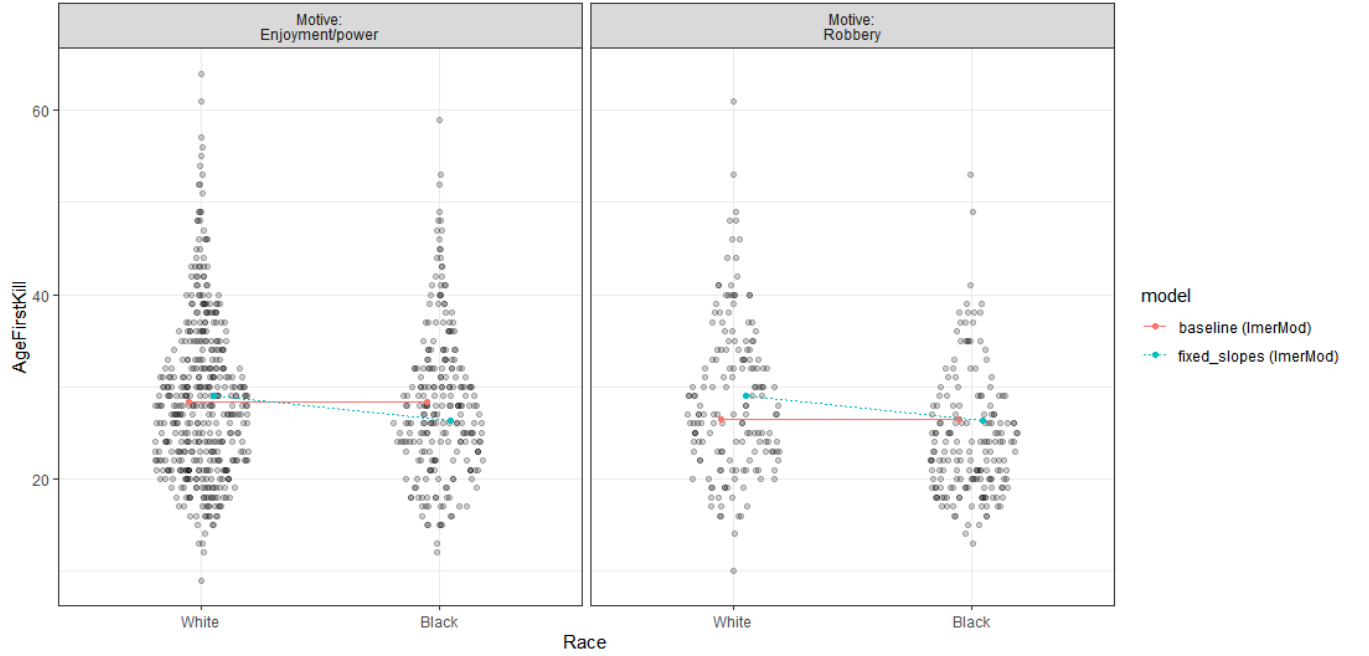


Figure 9: RIM Violin Plot - Balanced Covariate

Plot showing model (10) for the most prevalent motives *Enjoyment/power* and *Robbery*, equivalent to plots seen in figure (7), but with the covariate in question being **Race**. As seen in figure (9), a large sample is present between motives and between racial groups - issues of uncertainty now do not arise as discussed in the previous application. A natural continuation from a random intercept model is to consider relaxing the constraint of assuming the relationship between our response variable  $Y$  and feature input  $X$  for each cluster is equivalent i.e. to allow slopes to vary.

## 5 Random Slopes Model (RSM)

Here we introduce an extension to the level 2 RIM seen in section (4). Unlike the random intercepts model, the random slopes model (RSM) allows the possibility that the effect of a covariate on the response might also vary between clusters. A random slopes model applied to our sample  $\{y_{ij}\}$  is defined such that

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{1ij} + \sum_{l=2}^p \beta_l x_{lij} + e_{0ij}, \quad (11)$$

where all random terms on the right hand side are mutually independent and normally distributed,

$$e_{0ij} \sim N(0, \sigma_{e0}^2), u_{0j} \sim N(0, \sigma_{u0}^2) \text{ and } u_{1j} \sim N(0, \sigma_{u1}^2).$$

Whilst these random terms are assumed to be mutually independent, an exception is made for a possible covariance between the random effects,  $cov(u_{1j}, u_{0j}) = \sigma_{u01}$ . We can think of  $u_{1j}$  as the level 2 random effect of covariate  $x_1$  for group  $j$ . For the covariance between the random effects, a positive value indicates that clusters with high intercept residuals  $u_{0j}$  tend to have high slope residuals  $u_{1j}$ . Not only do we need to consider the sign of the covariance term, but the signs of the intercept  $\beta_0$  and slope  $\beta_1$  must be examined too.

For example, if both the slope and intercept are positive, a positive  $\sigma_{u01}$  suggests that clusters with large intercepts i.e., large  $\beta_0 + u_{0j}$ , on average, have steeper slopes- high  $\beta_1 + u_{1j}$ . The complement is true too- groups with low intercepts have flatter slopes than the average. This will result to a fanning out of the cluster respective regression lines when plotted together. The same reasoning can be made for negative values of  $\sigma_{u01}$ , with the outcome of regression line fanning inwards. We have  $p + 4$  unknown, constant parameters to estimate: fixed effects  $\beta_0, \dots, \beta_p$ , and variance parameters  $\sigma_{e0}^2$ ,  $\sigma_{u0}^2$  and  $\sigma_{u1}^2$ .

### 5.1 Application

Recall the following hypotheses that were proposed in section (4):

1. The distribution of the age at first kill varies within the race of a serial killer. Black ethnic groups tend to start killings at an earlier age compared to white serial killers.

In the previous section, we allowed for motive effects on the mean age a serial commits their first murder by allowing the intercept of the regression of **AgeFirstKill** on **Race** to vary across the motives. We assumed, however that changes in **AgeFirstKill** seen by **Race** are the same for all motives, i.e., the slope of the regression line were assumed fixed across motives. Using the random slopes model, we can consider the equivalent hypothesis proposed in the previous section, but instead of fixing the changes in **AgeFirstKill** seen by **Race** for all motives, we allow the intercept and slope to vary randomly across motives. Consider the model:

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{1ij} + e_{0ij}, \quad (12)$$

where  $y_{ij}$  denotes the **AgeFirstKill** of the  $i$ -th serial killer in the  $j$ -th cluster, **Motive**. This model is equivalent to (11) with only one covariate present  $x_{1ij}$ , the **Race** of the  $i$ -th serial killer in the  $j$ -th cluster.

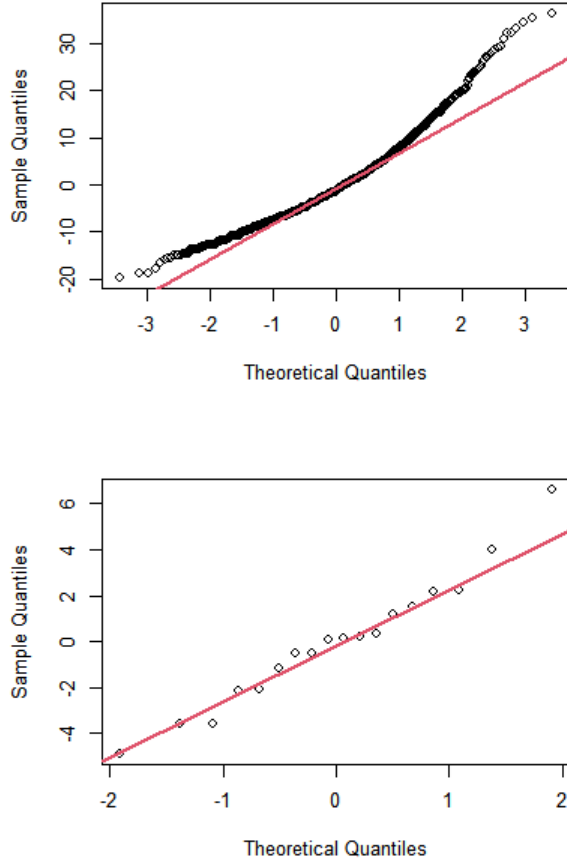


Figure 10: RSM Quantile-Quantile Plots Of Level 1 & 2 Residual Error

Top Left: Level 1 standard error residuals  $\{e_{0ij}\}$  quantile-quantile plot. Top Right: Level 2 random individual error  $\{u_{0j}\}$  quantile-quantile plot. 18 points are seen which are reference to the 18 unique values the variable **Motive** can take. Bottom Left: Level 2 random cluster error  $\{u_{1j}\}$  quantile-quantile plot. 18 points are again seen with the same reasoning as the random individual error case. Only a marginal difference is seen in the distribution of RSM residuals in comparison to VCM residuals - I again refer you to figure (4) for implications.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	25.39	24.03-26.75	<0.001
Race[White]	4.29	2.17-6.41	<0.001
<b>Random Effects</b>			
$\sigma_{e0}^2$	8.29 <sup>2</sup>		
$\sigma_{u0Motive}^2$	1.88 <sup>2</sup>		
$\sigma_{u1Motive.RaceWhite}^2$	3.39 <sup>2</sup>		
$\rho_{01Motive}$	0.15		
$ICC_{Intercept}$	0.04		
$N_{Motive}$	18		
Observations	1649		

Table 6: RSM Estimates For Model (12), Age-FirstKill  $\sim 1+Race + (1+Race|Motive)$

From figure (10), it seems justifiable to assume the underlying distribution of the residuals to be normal. Application of the random slopes model yields parameter estimates  $\hat{\beta}_0 = 25.39$  and  $\hat{\beta}_1 = 4.29$ . There is significant evidence to suggest black ethnic groups do tend to start killings at an earlier age compared to white serial killers as hypothesised- positive domained  $CI = [2.17, 6.41]$  and small p-value  $p < 0.001$  present in our findings. Whilst fixing the effect of **Race** across motives gave the parameter  $\hat{\beta}_1 = 2.63$  years in RIM, allowing variation within motives in this random slopes model has increased the impact of our feature input substantially to  $\hat{\beta}_1 = 4.29$  years.

The intercept-slope correlation is defined as

$$\rho_{u01} = \frac{\sigma_{u01}}{\sqrt{\sigma_{u0}^2 \sigma_{u1}^2}} = \frac{0.951}{\sqrt{1.88^2 3.39^2}} = 0.15, \text{ for the example above.}$$

In the intercept-slope correlation definition, we consider the term  $\hat{\sigma}_{u01} = 0.951$ , the covariance between the intercept and slope at the motive level (the covariance of the random effects). This takes a large positive value which indicates a strong positive relationship between the random effects at the cluster-level. The useful way to interpret such a result is to consider the fixed effects ( $\beta_0$  and  $\beta_1$ ) coupled with the covariance  $\sigma_{u01}$ . From table (6) we find both the slope, intercept and covariance  $\hat{\sigma}_{u01} = 0.951$  take positive values. This suggests that groups with large intercepts i.e., large  $\beta_0 + u_{0j}$ , on average, have steeper slopes- high  $\beta_1 + u_{1j}$ . White serial killers, on average, tend to start killings at an older age. Vice versa black serial killers, on average, tend to start killings at a younger age.

Whilst considering the numerator (thus the sign) of the intercept-slope correlation  $\rho_{u01}$  at the cluster-level relative to the fixed effects has provided insight between the race and age at first kill of a serial killer, considering the absolute value of the intercept-slope correlation should be considered too. In doing so will help us understand whether variations between killers with different motives that has been shown to be apparent (see previous section) is mostly due to our predictor **Race**, or mostly due to inherent differences between serial killers and the age at which they commit their first murder. A weak correlation is present between the intercept and slope random effects at the motive level-  $\rho_{01\text{Motive}} = 0.15$  with such a value spanning between 0 (no correlation) and 1 (direct correspondence between intercept and slope) - to some extent, this enables us to lax our dependency assumption between  $u_{0j} \sim N(0, \sigma_{u0}^2)$  and  $u_{1j} \sim N(0, \sigma_{u1}^2)$ . Note: whilst small correlation between effects does not necessarily imply independence (or a small amount of dependency), the fact that our random effects  $u_{0j}$  and  $u_{1j}$  both follow equivalent distributions enables us to make such a statement which will be used as follows.

Intraclass correlation coefficients cannot be applied to random slope models due to the increase in complexity given by the level 2 random effect  $u_{1j}$  [8] however, given a “small” amount of dependency present between level 2 random effects  $u_{0j}$  and  $u_{1j}$  as evident by the “small” absolute value of the intercept-slope correlation  $\rho_{01\text{Motive}} = 0.15$  it is useful to define the statistic

$$ICC_{\text{Intercept}} = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{u1}^2 + \sigma_{e0}^2} = \frac{1.88^2}{1.88^2 + 3.39^2 + 8.29^2} = 0.04, \text{ as given in the table above.} \quad (13)$$

This is a measure of the variability seen at the intercept in relation to the total level 1 and 2 variance  $\sigma_{u0}^2 + \sigma_{u1}^2 + \sigma_{e0}^2$ . Whilst  $ICC_{\text{Intercept}}$  is in essence, not a correlation metric, keeping intraclass correlation coefficient initials is common practice seen in multilevel modeling literature- see J Hox (et al) 2017, “Multilevel analysis: Techniques and applications” for an example of its use. It seems only a small amount of variability seen within serial killers age at first kill and race are explained at the motive level.

## 5.2 Limitations

Relaxing the constraint of assuming the relationship between our response variable  $Y$  and feature input  $X$  for each group has added an extra layer of complexity in the random slopes model in comparison to the random intercepts model. Is this complexity justified? From the example prior regarding how the distribution of the age at first kill varies within the race of a serial killer, we have seen a substantial increase in relation between race and age at first kill from the RIM with race parameter  $\hat{\beta}_1 = 2.63$ , to the more complex RSM taking value  $\hat{\beta}_1 = 4.29$  years. However, variation between motives present due to the age at which a serial killer commits their first murder seems to have fallen with intraclass correlation coefficient values of  $ICC = 0.13$  and  $ICC_{\text{Intercept}} = 0.04$  respectively. Whilst the latter statistic, the  $ICC_{\text{Intercept}}$ , is an adjusted version of the intraclass correlation coefficient as explained above, it still remains useful to compare both statistics [12]. Less variability in age at first kill is seen at the motive level. Is multilevel modeling justified to answer the hypothesis given in section (4.3) and (5.1), and if so, which presented model RIM or RSM is best suited to answer our hypothesis?

## 6 Model Comparison

To determine whether multilevel modeling is justified to answer given hypotheses about serial killers, we should ask the obvious question that is, are the additional random terms really necessary? There might be some common sense reasons for including the random terms as mentioned during our model application examples for instance, the belief that the age at first kill of serial killers varies within motives nevertheless, we can also ask if the observed data supports this from an objective standpoint. So our focus becomes under which of these models, the GLM applied in section (2) against their respective multilevel model extensions: VCM, RIM or RSM is our observed sample “most likely”<sup>4</sup> to have occurred?

### 6.1 Likelihood Ratio Test

The likelihood ratio test can be used to compare nested models that is, the terms present in one of the two compared models in question is a subset of the latter. This test is useful in determining if whether the addition of a feature input is useful in answering the hypothesis in question as compared to the exclusion of the feature. To determine the likelihood of our data under a given model, consider the joint probability density of our sample under that model,  $f(y_{11}, y_{21}, \dots, y_{n_{mm}})$ , evaluated at the observed values of  $y_{11}, y_{21}, \dots, y_{n_{mm}}$ , where  $n_m$  denotes the number of individuals situated in cluster  $j = m$ . The greater this density value is the more “likely” these observations are under that model. However, when computed at the observed values,  $f()$  is no longer thought of as a function of the sample, instead, it is a function of the unknown model parameters. For example in a random intercepts model, the likelihood is defined as

$$L(\beta_0, \dots, \beta_p, \sigma_{e0}^2, \sigma_{u0}^2) = f(y_{11}, y_{21}, \dots, y_{n_{mm}}), \quad (14)$$

where our aim would be to maximise the likelihood subject to varying these unknown model parameters  $\beta_0, \dots, \beta_p, \sigma_{e0}^2, \sigma_{u0}^2$ , with fixed values that accomplish such a maximisation taking the values  $\hat{\beta}_0, \dots, \hat{\beta}_p, \hat{\sigma}_{e0}^2, \hat{\sigma}_{u0}^2$  respectively. These values best explain the data under the framework of the model given and are formally referred to as maximum likelihood estimates (MLEs) of the parameters. As mentioned in section (3.1), multilevel modeling uses a slightly more complicated parameter derivation method that is, iterative generalised least squares however, such a method converges to maximum likelihood methods and so applying likelihood ratio tests to multilevel models is justified [14]. Define our maximised likelihood estimate as

$$\hat{L} = L(\hat{\beta}_0, \dots, \hat{\beta}_p, \hat{\sigma}_{e0}^2, \hat{\sigma}_{u0}^2), \text{ for the random intercept model case.} \quad (15)$$

Then, to compare two nested models, A and B, where A is nested within B, we can consider their maximised likelihoods  $\hat{L}_A$  and  $\hat{L}_B$  by taking the likelihood ratio  $\hat{L}_B/\hat{L}_A$ . If such a ratio takes a value close to 1, then there is no meaningful reason to prefer the model B over A because choosing such a model does not increase the likelihood of the data by a substantial amount. Alternatively, if the ratio is “large” then it seems that the likelihood of the data is greater under B. A measurement of such model disparity can be calculated via the deviance statistic.

#### 6.1.1 Deviance Statistic

The deviance statistic is defined as

$$D = 2 \log \frac{\hat{L}_B}{\hat{L}_A}, \text{ where } \hat{L}_A \text{ and } \hat{L}_B \text{ are the maximum likelihood values of models A and B respectively.} \quad (16)$$

Under the assumption that model A is correct we can assume  $D \sim \text{approx } \chi_d^2$ , where  $d$  denotes the number of additional parameters introduced in model B as compared to A [14]. The larger the deviance statistic of our model comparison, the greater evidence in favour of preferring B over A. We test the the hypotheses:

<sup>4</sup>By comparing models and choosing the “most likely” model given serial killer data, it is important to consider model complexity in conjunction to how well the data is explained. This is applied and explained in our model comparison tests.



$$\begin{cases} H_0 : & \text{All additional parameters in model B that are not in A are zero, against the alternative} \\ H_1 : & \text{at least one of the additional parametrs in model B is nonzero.} \end{cases}$$

Basing our test on the significance level  $\alpha = 5\%$ , as standard. A test in which we reject  $H_0$  if and only if

$$D \geq \chi_d^2(\alpha) \quad (17)$$

is called a **likelihood ratio test** with significance level  $\alpha$  for model B against that of model A. Note,  $\chi_d^2(\alpha)$  is defined as the  $(1 - \alpha)$  quantile of the  $\chi_d^2$  distribution i.e., the 95% quantile of the chi-squared distribution with  $d$  degrees of freedom.  $\chi_d^2(\alpha)$  is the number such that the  $P(D \geq \chi_d^2(\alpha)) = \alpha$ . It is also worth noting that the deviance statistic does not consider the likelihood of the models in question but the log-likelihood  $D = 2 \log(\hat{L}_B / \hat{L}_A) = 2 [\log(\hat{L}_B) - \log(\hat{L}_A)]$ . Reasoning for this comes from the logarithm being a monotonically increasing function- parameter values that maximise the likelihood will indeed maximise the log-likelihood. Furthermore, computational efficiencies are brought to the table when considering a logarithm-transformed function in comparison to applying computation to the raw values themselves [15].

### 6.1.2 Application

<i>Model A</i>		<i>Model B</i>		
AgeFirstKill ~ 1		AgeFirstKill ~ 1 + (1 Motive)		
AgeFirstKill ~ 1+Sex		AgeFirstKill ~ 1+Sex + (1 Motive)		
AgeFirstKill ~ 1+Race		AgeFirstKill ~ 1+Race + (1 Motive)		
AgeFirstKill ~ 1+Race		AgeFirstKill ~ 1+Race + (1+Race Motive)		
AgeFirstKill ~ 1+Race + (1 Motive)		AgeFirstKill ~ 1+Race + (1+Race Motive)		
<i>Likelihood <math>\log \hat{L}_A</math></i>	<i>Likelihood <math>\log \hat{L}_B</math></i>	<i>Deviance Statistic <math>D</math></i>	<i>Quantile <math>\chi_d^2(\alpha = 5\%)</math></i>	<i>Null Hypothesis <math>H_0</math></i>
-6308.093	-6272.595	70.997	5.024	reject
-6304.245	-6272.426	63.637	5.024	reject
-5893.685	-5862.231	62.909	5.024	reject
-5893.685	-5849.818	87.734	7.378	reject
-5862.231	-5849.818	24.825	5.024	reject

Table 7: Likelihood Ratio Test Statistics For GLM, VCM, RIM & RSM

Table (7) shows the generalized linear models, variance components model, random intercept and random slopes models proposed and analysed in sections (2-5) respectively. The simpler model i.e., the model with fewer parameters is proposed as model A, and consists as a nested structure within the proposed, more complex model B. For example, the first row of table (7) compares the most simple GLM with no feature inputs **AgeFirstKill** ~ 1, to that of the multilevel model extension, the VCM given by **AgeFirstKill** ~ 1 + (1|Motive). As a reminder, meanings behind this formula notation are given in appendix (A.1). Subsequently, the GLM with the inclusion of a feature that is, **Sex**, is compared to its multilevel counterpart, the RIM given by **AgeFirstKill** ~ 1+Sex + (1|Motive).

Further model comparison are given, with the latter comparing whether the inclusion of the possibility that the effect of a covariate on the response might also vary between clusters provides greater “likelihood” in our findings- RIM vs RSM given by **AgeFirstKill** ~ 1+Race + (1|Motive) against **AgeFirstKill** ~ 1+Race + (1+Race|Motive). Likelihood ratio test statistics are seen for each model comparison. The maximum log-likelihood estimated value of the models  $\log \hat{L}_A$  and  $\log \hat{L}_B$ , the deviance statistic  $D$ , the quantile threshold  $\chi_d^2(\alpha = 5\%)$  and the result of the hypotheses proposed

during a likelihood ratio test- see section (6.1.1). For example, the deviance statistic in the first row of table (7) is calculated using the definition provided in section (6.1.1) giving:

$$\begin{aligned}\text{Deviance Statistic } D &= 2 \left[ \log \hat{L}_{A=\text{GLM Base Case}} - \log \hat{L}_{B=\text{VCM}} \right] \\ &= 2 [-6308.093 + 6272.595] \\ &= 70.997.\end{aligned}$$

All quantile thresholds  $\chi_d^2(\alpha = 5\%)$  take equivalent values that of  $\chi_{d=1}^2(\alpha = 5\%) = 5.024$  except for the penultimate model **AgeFirstKill ~ 1+Race** against **AgeFirstKill ~ 1+Race + (1+Race|Motive)**, taking quantile threshold value  $\chi_{d=2}^2(\alpha = 5\%) = 7.378$ . This is because all model comparison seen only consider an increase in complexity of one parameter, resulting in a net difference in degrees of freedom between model A and B as  $d = 1$ - the penultimate model jumps up by two. Throughout all model comparisons we reject the null hypothesis  $H_0$  that is, all additional parameters in model B that are not in A are zero<sup>5</sup>. Since our deviance statistic  $D \geq \chi_d^2(\alpha)$  we can say that, under the assumption that model A is correct (i.e. the null hypothesis that there is no random serial killer effect) our deviance statistic is improbably large- less than 5% chance of being so large. Hence observing such a large statistic is strong enough evidence against our null hypothesis  $H_0$  that we can reject it at the 5% significance level, in favour of the hypothesis that model B is preferred.

Thus we may conclude from likelihood ratio testing that multilevel modeling is indeed justified to answer given hypotheses about serial killers; additional random terms really can be deemed as necessary and provide more insight, more “likelihood”, in our findings. Likelihood ratio testing is one of many tools that could be used to compare multilevel models. More comparison statistics should be considered in order to justify one model over another.

## 6.2 Further Model Comparison Metrics

So far we have focused on a maximum likelihood based model selection, a measure of how well the model is able to fit to the available data, when the parameters are chosen to make this fit as good as possible. A possible issue with this approach is that with more parameters we can almost always tweak the values to make the likelihood just a little higher. For instance, consider the RSM given by **AgeFirstKill ~ 1+Race + (1+Race|Motive)**:

$$y_{ij} = (\beta_0 + u_{0j}) + (\beta_1 + u_{1j})x_{1ij} + e_{0ij}, \quad (18)$$

where  $y_{ij}$  denotes the **AgeFirstKill** of the  $i$ -th serial killer in the  $j$ -th cluster, **Motive**. During maximum likelihood based model selection such as the likelihood ratio test, we vary the parameters  $\hat{\beta}_0, \dots, \hat{\beta}_p, \hat{\sigma}_{e0}^2, \hat{\sigma}_{u0}^2, \hat{\sigma}_{u1}^2$  with the goal to maximise the “likelihood” of the observations occurring  $\hat{L} = L(\hat{\beta}_0, \dots, \hat{\beta}_p, \hat{\sigma}_{e0}^2, \hat{\sigma}_{u0}^2, \hat{\sigma}_{u1}^2)$ . Consider the RIM given by **AgeFirstKill ~ 1+Race + (1|Motive)**:

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{1ij} + e_{0ij}, \quad (19)$$

where  $y_{ij}$  denotes the **AgeFirstKill** of the  $i$ -th serial killer in the  $j$ -th cluster, **Motive**. The RSM is equivalent to that of RIM with the inclusion of some perturbation in our covariate term via  $u_{1j} \sim N(0, \sigma_{u1}^2)$ . We can always make the RSM consist of a greater (or at least equivalent) maximum likelihood just by first setting the level 2 residual  $\hat{\sigma}_{u1}^2$  to zero and varying the value such that the likelihood increases. In table (7), the fit for the RSM model is only marginally improved, in terms of likelihood, than the RIM with maximum log-likelihood values of  $-5862.231$  and  $-5849.818$  respectively. We might, qualitatively, think “that little bit of extra likelihood isn’t worth having an extra parameter”. How do we make that idea quantitative? One way is to apply a hypothesis test as seen in the likelihood ratio method.

<sup>5</sup>New findings seen in Goldstein’s 4<sup>th</sup> edition of Multilevel Statistical Models (page 29) deem our quantile threshold  $\chi_d^2(\alpha = 5\%)$  of being too strict. Fortunately, this discovery does not change our results- we would still reject the null hypothesis under the new quantile value. Halving the p-value of the hypothesis test i.e., considering the threshold  $\chi_d^2(\alpha = 10\%)$  should be used for future work. Reasoning stems from the test statistic being a weighted mixture of chi-squared distributions.

Another is to use a penalised likelihood that explicitly “punishes” the extra parameters. Usually this takes the form of an extra term, in the log-likelihood, proportional to the number of parameters  $p$ , with a penalisation strength  $\lambda$ :

$$\log L_{\text{PENALISED}} = \log \hat{L} - \lambda p, \quad (20)$$

where  $\log \hat{L}$  is the usual maximum log-likelihood. Unfortunately there is no absolute consensus in the statistical community on how to choose  $\lambda$  [7]. It may depend on your personal experience or needs regarding model complexity.

### 6.2.1 Akaike Information Criterion (AIC)

Probably the most prevalent penalised likelihood is known as the **Akaike Information Criterion (AIC)**. This is defined as:

$$AIC = 2p - 2\log \hat{L}, \quad (21)$$

effectively setting  $\lambda$ , seen in equation (20), to be equal to one, and the model that minimises the AIC is deemed to be the best [7]. Setting  $\lambda$  equal to one makes equation (19) proportional to the AIC definition. It is proportional due to the fact the AIC is not just some arbitrary formula. Instead the Akaike information criterion is derived from minimizing the Kullback–Leibler divergence, a type of “statistical distance” that measures how one probability distribution is different from another. Further information can be found in Hox J’s (et al) Handbook of Advanced Multilevel Analysis [7].

### 6.2.2 Bayesian Information Criterion (BIC)

Similar to AIC, the **Bayesian Information Criterion (BIC)** can be used as a measure regarding model selection, penalising the likelihood of the models not by setting  $\lambda$  equal to one as seen by the AIC metric, instead,  $\lambda = \log(n)/2$  with  $n$  denoting the number of observations of serial killers present hence:

$$BIC = p\log(n) - 2\log \hat{L}, \quad (22)$$

where  $\log \hat{L}$  is the usual maximum log-likelihood and  $p$  the number of parameters. Unlike the Akaike information criterion that derives from a frequentist approach to comparing models, BIC is a metric derived from a Bayesian approach, where prior beliefs of the model in question are taken into account [7]. A “flat prior” is considered for this penalised likelihood metric. This is a specific type of prior distribution subject to the distribution of the current model in question. Its meaning and derivation is not unique to multilevel analysis and can be explored by the reader through work by N. Kumar (et al) 2010, “On the derivation of the Bayesian Information Criterion”.

### 6.2.3 Application

<i>Model</i>	<i>AIC</i>	<i>AIC Rank</i>	<i>BIC</i>	<i>BIC Rank</i>
AgeFirstKill ~ 1	12620.19	8	12631.14	8
AgeFirstKill ~ 1 + (1 Motive)	12551.19	6	12567.61	4
AgeFirstKill ~ 1+Sex	12614.49	7	12630.91	7
AgeFirstKill ~ 1+Sex + (1 Motive)	12552.85	5	12574.75	5
AgeFirstKill ~ 1+Sex + (1+Sex Motive)	12548.73	4	12581.58	6
AgeFirstKill ~ 1+Race	11793.37	3	11809.59	3
AgeFirstKill ~ 1+Race + (1 Motive)	11732.46	2	11754.09	2
AgeFirstKill ~ 1+Race + (1+Race Motive)	11711.64	1	11744.08	1

Table 8: Further Model Comparison Metrics For GLM, VCM, RIM & RSM

As seen in table (8), model comparison metrics may be applied to our previously seen examples. The table shows the AIC and BIC values for the multilevel models and the generalized models seen in section (2.1). Here, we will also discuss ICC values for specific multilevel model with the meaning behind intraclass correlation coefficient values being found in section (3.1) with an example of computation given. As a reminder, ICC values have been left regarding generalized linear models because the intraclass correlation coefficient is a measure of variability between individuals at a cluster level, thus making it not applicable in the GLM case. For RSM, an adjusted ICC value will be discussed and is defined by equation (13). Firstly however, an example of how the AIC value for the first model, **AgeFirstKill** ~ **1** is as follows:

$$\begin{aligned} AIC_{\text{GLM Base Case}} &= 2p_{\text{GLM Base Case}} - 2\log \hat{L}_{\text{GLM Base Case}}, \\ &= 2(2) - 2(-6308.093), \\ &= 12620.19. \end{aligned} \tag{23}$$

A ranking between all cases can be seen taking values from 1, the model produces the most optimum performance metric relative to other case to 8, the model performs least. This ranking format can be used to compare desired cases for example, we may ask the following question:

- Does the effect of **Race** vary between serial killers clustered by **Motive**?

Whilst the previous model comparison method, the likelihood ratio test, suggested a random effect to be most applicable, i.e., the RSM **AgeFirstKill** ~ **1+Race** + (**1+Race|Motive**) has a greater preference at explaining the data observed in comparison to the RIM **AgeFirstKill** ~ **1+Race** + (**1|Motive**); metrics found in table (8) suggest otherwise. Whilst the RSM does indeed have greater performing AIC and BIC values in comparison to the RIM, these differences are only marginal- (RIM, RSM) with values (11732.46, 11711.64) respectively for the Akaike information criterion, and (RIM, RSM) with values (11754.09, 11744.08) for the Bayesian information criterion. A greater difference is prevalent regarding intraclass correlation coefficient values (RIM, RSM) with (0.132, 0.042) suggesting that the majority of variance explained is seen from the fixed effect, not the random effect of the models [14].

### 6.3 Treating Motive As A Fixed Effect

Throughout our model application examples, we have considered **Motive** only at the level 2, cluster level. A natural question arises by doing so that is, does considering **Motive** at the cluster level bring more insight into understanding the criminal mind than considering the motive of a serial killer as a covariate that is, at the level 1 individual serial killer level? In other words, which model is of greater use at explaining serial killers- clustering motive with **AgeFirstKill** ~ **1** + (**1|Motive**) as analysed prior, or regressing motive **AgeFirstKill** ~ **1+Motive**. We introduce the model:

$$\text{AgeFirstKill}_i = \beta_0 + \beta_1 \text{Motive}_i + e_i, \tag{24}$$

where  $\text{Motive}_i$  and  $\text{AgeFirstKill}_i$  refer to the motive and age at first kill of the  $i$ -th serial killer,  $\beta_0$  and  $\beta_1$  unknown parameter constants and  $e_i$  the residual of the  $i$ -th serial killer, following a normal distribution. Since **Motive** is a categorical variable, taking 18 unique values we partition this covariate into 17 separate inputs or dummy variables, one for each motive with encoding “1”, the serial killer does indeed take the specific motive in question and encoding “0”, the killer does not. We require 17 dummy variables not 18 as may be first thought since a serial killer must take at least one motive i.e., all covariates must not take the value zero. Even if the motive of a serial killer is not known- this scenario in itself has its own category, its own dummy variable. We pivot around the motive *Angel of Death*, arbitrary, internalizing this motive within our intercept parameter  $\beta_0$ . Thus we may rewrite equation (24) as:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_{17} x_{17i} + e_i, \tag{25}$$

where  $y_i$  denotes the **AgeFirstKill** of the  $i$ -th serial killer and  $x_{1i}, \dots, x_{17i}$  refer to the 17 unique serial killer motives with equivalent order of labels to that seen in figure (3) i.e.,  $j = 1$  refers to the motive *Anger*,  $j = 2$  the motive *Avoid*

*arrest* and so forth up to  $j = 17$ , *Unknown* motive. We may now estimate parameters  $\beta_0, \dots, \beta_{17}$  via minimizing the ordinary least squares of the residuals producing the following results.

AgeFirstKill			
Predictors	Estimates	CI	p
(Intercept)	29.84	26.88 – 32.80	<b>&lt;0.001</b>
Anger	-3.01	-6.25 – 0.24	0.069
Avoid arrest	2.11	-2.78 – 6.99	0.398
Black widow	2.53	-1.46 – 6.52	0.214
Convenience	2.76	-3.23 – 8.75	0.366
Cult	-8.15	-13.59 – -2.70	<b>0.003</b>
Enjoyment/power	-1.45	-4.47 – 1.57	0.346
Financial/personal gain	5.72	1.96 – 9.49	<b>0.003</b>
Mental illness	-2.95	-9.19 – 3.29	0.354
Missionary	-1.07	-6.05 – 3.90	0.672
Multiple motives	-5.81	-9.13 – -2.50	<b>0.001</b>
Munchausen syndrome	-6.42	-12.02 – -0.82	<b>0.025</b>
Organised crime	-8.08	-12.51 – -3.65	<b>&lt;0.001</b>
Prejudice racism	-3.10	-7.43 – 1.24	0.161
Revenge/vigilantism	-3.38	-7.06 – 0.29	0.071
Robbery	-3.37	-6.45 – -0.28	<b>0.032</b>
Sniper	0.99	-6.35 – 8.34	0.791
Unknown	-2.04	-5.54 – 1.45	0.252
Observations	1763		
R <sup>2</sup>	0.068		
AIC	12529.37		
BIC	12633.39		

Table (9) shows the estimates of model parameters  $\hat{\beta}_0 = 29.84$ , the intercept,  $\hat{\beta}_1 = -3.01$ , the effect of motive *Anger*, up to parameter estimate  $\hat{\beta}_{17} = -2.04$ , *unknown* motive. Interpretation of the coefficient values is as follows. The sign of the coefficients determines the effect of that motive on the age a serial killer commits their first murder with respect to the motive *Angel of Death*, the base case internalised in the intercept parameter- negative sign implies the motive to constitute killers from a younger demographic, positive sign older demographic. The absolute value of the parameter estimates indicates the dominance of that effect. For instance, *Cult* - including *human sacrifice* seems to comprise of serial killers with ages at first kill, on average, starting at 8 years younger compared to the base case- *Financial/personal gain* around 6 years older. The significance of each motive inference can be judged by confidence intervals and p-values seen in the table, the meaning behind such metrics are given in earlier analysis in section (2).

Table 9: GLM Estimates For Model (25), AgeFirstKill ~ 1+Motive

### 6.3.1 Limitations

Treating **Motive** as a fixed effect may be the first thought of statisticians trying to find relations between the motive and age at which a serial killer commits their first murder however, doing so comes with its limitations. Yes, a significant amount of variation in the age a serial commits their first murder can be seen by analyzing **Motive** as a fixed effect as evident by some fairly large, and significant, motive parameter values discussed above however, this approach does not allow for some useful metrics that can quantify just how much variation is present at the motive level as a multilevel modeling approach would. Intraclass correlation comparisons cannot be made when motive is treated as a covariate.

Whilst considering motive as a covariate is an important case to consider, in doing so we may be over-simplifying our model. Relaxing the assumption that our level 2 random error  $u_{0j} \sim N(0, \sigma_{u0}^2)$  could be argued to be under-fitting the data [6]. The standard regression model is too simple to describe serial killers and so upon application results in poor parameter estimates, poor explanation of the data relative to the true underlying patterns seen. Furthermore, treating **Motive** as a fixed effect forces us to internalise one of the 18 unique motives as serial killer can take into the model framework, treating the chosen motive as a base case. In the example above, we chose the base case motive *Angel of Death*, arbitrary. It seems that this context has no intuitive base case to choose- there is no single motive that you would naturally pick and compare against other motives. This would question whether this standard fixed effect approach is justifiable in this context- multilevel modeling seems to be a better choice here.

I also think that one of the strengths of multilevel modeling in regards to understanding the criminal mind is the simplicity of the random slope model. It allows us to get a measure of how much the effect of race varies between

motives by estimating a single parameter (the level 2 variance of  $u_{0j}$ ). As seen, a similar analysis with a fixed effect model has required us to fit a model with many more parameters (an interaction effect for race with 17 different motives), leading to a much more complex and less interpretable model. If the question is simply "Does the effect of race vary by motive and, if so, by how much?" then the multilevel approach seems quite attractive.

## 6.4 Summary

In this section we have explored multiple ways of comparing models to each other, explaining the rationale behind such methods and applying them to our previously analysed models. Likelihood ratio testing, comparing Akaike information criterion and comparing Bayesian information criterion have been the quantitative approaches discussed that have the ability to compare model performances. Using these quantitative results coupled with qualitative judgments made by us, the statisticians, there are three cases explored that we may determine the most preferable model. Recall the three questions asked throughout our multilevel modeling journey.

1. How does the age of a serial killer vary when they commit their first murder?
2. Does the distribution of the age at first kill depend on the gender of the serial killer?
3. The distribution of the age at first kill varies within the race of a serial killer. Black ethnic groups tend to start killings at an earlier age compared to white serial killers.

The model that is preferred to answer our first question regarding variation between the age a serial killer commits their first murder is **AgeFirstKill**  $\sim 1 + (1|\text{Motive})$ , i.e., the variance components model. The alternative models that could be used to answer the first question are **AgeFirstKill**  $\sim 1$ , a standard regression, and **AgeFirstKill**  $\sim 1 + \text{Motive}$ , regressing motive as a fixed effect. This winning model could be argued to describe the data best with the lowest BIC value- 12567.61 compared to 12631.14 and 12633.39 seen from the other two model respectively (see table (8)). Whilst there is some disagreement seen by AIC values- 12551.19 compared to 12620.19 and 12529.36 seen from the alternatives, the variance components model is still preferred to answer our first question. This is because qualitatively speaking, I would argue despite regressing motive as a fixed effect resulting in the model with the lowest AIC value, applying such a model to answer our question regarding variation between killers is undesirable. As discussed in section (6.3.1), a fixed effect model has required us to fit a model with many more parameters with interaction effects for race with 17 different motives, leading to a much more complex and less interpretable findings. Due to small serial killers samples present in certain motives, the fixed effect approach could be deemed to be over-fitting the data too- internalizing random noise seen in the data in the parameter estimates. Natural clusters do indeed seem to form between serial killers with different motives- the variance components model is ideal for such a scenario.

For the second question we considered a covariate term to find relations between the gender of serial killers and their age at first kill. In my opinion, the most preferable model from the models explored that could best answer the second question is **AgeFirstKill**  $\sim 1 + \text{Sex}$  i.e., the standard regression model. The alternative models explored were **AgeFirstKill**  $\sim 1 + \text{Sex} + (1|\text{Motive})$ , the random intercept model, and **AgeFirstKill**  $\sim 1 + \text{Sex} + (1 + \text{Sex}|\text{Motive})$ , the random slopes model. As discussed in previous sections, issues from applying these alternative multilevel models stem from the large imbalance in the number of female and male serial killers. Only a small proportion of serial killers are female (around 8% of the total population), thus partitioning our data into multiple smaller subsets (clustering by motives) and regressing within these subsets resulted in insignificant findings as discussed in section (4.2). Standard regression does not succumb to such a limitation and answers the second question clearly, especially since no focus on variability between serial killers is mentioned in our second question.

The model that is most preferable to answer our third and final question exploring variations between serial killer races and their age at first kill is deemed to be the random intercept model that is, **AgeFirstKill**  $\sim 1 + \text{Race} + (1|\text{Motive})$  - alternative models are standard regression: **AgeFirstKill**  $\sim 1 + \text{Race}$ , and the random slopes model: **AgeFirstKill**  $\sim 1 + \text{Race} + (1 + \text{Race}|\text{Motive})$ . Justification for such a model preference takes into account both the quantitative metrics discussed in this section and the qualitative findings seen by us, the statisticians. We first disregard the standard regression model due to the question above discussing variability between races to which standard regression does not answer in a satisfactory manner- intraclass correlation coefficients and variation parameters are not present in such a

model as discussed in previous sections. In regards to choosing between the random intercept or random slope models; whilst our quantitative analysis applied in this section suggests the random slope model to be the most attractive out of the two models with the lowest AIC and BIC values coupled with a favourable likelihood ratio test compared to the random intercept model, I would still choose the random intercept model over the random slopes model. A discussion for this model preference is seen at the end of section (6.2) whereby I directly compare the two models. In short, when constructing a model with the goal of answering a question it is important produce a model that is complex enough to answer the question in a satisfactory manner but no more complex than that. As Einstein once famously said, “Everything should be made as simple as possible, but not simpler”.

## 7 Multivariate Multilevel Models & Profiling Serial Killers

Criminology is a vital part of law enforcement and public safety in the modern age. Action and public policy made according to the findings of studies in this field have proven effective and popular. However, efforts to reduce and understand serial killings haven't received much of this windfall [11]. Perhaps by using a **multivariate multilevel modeling** (MMM) approach, we can close the gap between the qualitative and quantitative research on profiling killers, predicting their behaviour.

Multivariate response data are conveniently incorporated into a multilevel model by creating an extra level "below" the original level 1 units to define the multivariate structure [17]. We thus have responses within individuals that are in turn nested within higher-level units. There are many examples of data where multivariate multilevel modeling is widely used for instance, MMM has been applied to explore variations in education with interesting and useful results. The multilevel aspects stem from students (level 1) being taught in different schools (level 2), and the multivariate structure ("level 0") could be the marks of the students on various standardized exams or the gender of the students, or both. The medical industry is another good example that has seen benefits from multivariate multilevel modeling [18]. Patients could be clustered by hospitals with patient data such as symptom severity and survival rate treated as the "level 0", multivariate responses. In the context of serial killers, our main discussion will be trying to determine:

- Can multivariate multilevel modeling be used to profile serial killers, predicting their characteristics and behaviour?

Analogous to the examples of MMM given above, our multivariate multilevel models applied to serial killers will have the following general structure. We will be interested in using crime-scene data to profile killers by **Race**, **Sex** and **Motive**. These will be the "level 0", multivariate responses with the individual serial killers constituting level 1. For the level 2, cluster level case, we will be interested in variations of responses in different U.S. states- given by the aptly named variable **State** in the Radford database, a discrete variable describing the states in which a killer has committed their murders with state initials given. For example, a serial killer who killed in Los Angeles would have **State** output *LA*, California *CA* and a killer who murdered in both, *LA / CA*. Clustering by state seems like a good choice here that could yield fruitful results. For instance, the United States has a known history a racial disparity, with the northern central states such as Montana and West Virginia comprising of greater white/Caucasian groups compared to other states [13]. Given these natural clusters of races within states and the fact that serial killers tend to kill within their own communities, killing people who they know, you would expect to see some variations of **Race** at the state level. Furthermore, recall that the serial killer data consists of U.S. killers from the 1900s on-wards and the fact that apartheid- the application of racial segregation policies was only lifted in 1964, spanning more than half the time-frame in which data was gathered- this should supplement variations in racial demographics that could be seen within the data.

Killer $i$	State $j$	Response	Race	Sex	Motive
1	1	$y_{i=1j=1}^{(0)}$	1	0	0
1	1	$y_{i=1j=1}^{(1)}$	0	1	0
1	1	$y_{i=1j=1}^{(2)}$	0	0	1
5	2	$y_{i=5j=2}^{(1)}$	0	1	0
5	2	$y_{i=5j=2}^{(2)}$	0	0	1
21	10	$y_{i=21j=10}^{(1)}$	0	1	0

Table 10: Data Matrix Example For Multivariate Case

As compared to the univariate case whereby we consider a response  $y_{ij}$ , the  $i$ -th response of the  $j$ -th cluster; the multivariate framework will denote responses in a similar fashion, however with superscripts referencing the category of response we are considering. Table (10) is an example of such notation showing data on three serial killers, killers  $i = 1, 5$  and  $21$ , the state in which they committed their murders, states  $j = 1, 2$  and  $10$ ; as well as their response output  $y_{ij}$  for each category- **Race**, **Sex** and **Motive**. Not only does this example give us an idea of a multivariate structure, it also shows missing information for the latter two killers, with one missing information on the race ( $y^{(0)}$ ) of the killer and other the race and motive ( $y^{(0)}$  and  $y^{(2)}$ ).



In standard univariate cases, missing response values are usually removed from the model (or imputed depending on the type of missing data it resides in i.e. missing at random, MAR, missing completely at random, MCAR etc). However in the multivariate case, the relation seen between responses may aid in imputation of missing value cases [6]. For example, consider observing 10 serial killer race, sex and motives in relation to some feature input say, whether the murderer killed with a gun or not. In these observations, we find black, male killers linked to organised crime to be highly correlated with our predictor variable. Now consider observing a new serial killer yet unfortunately only know the killers sex and motive (criminal records have a tendency to be inconsistent and incomplete, especially those from multiple decades ago that reside in unorganized states, and when conflict such as civil wars occurred). If the killer's sex and motive was seen to comprise of a male killer in organised crime, then given these 10 prior observations, we are more inclined to suspect the serial killer to be of black ethnicity, aiding in imputation of this missing value. This Bayesian methodology is internalised within the parameter estimation process for the multivariate case [6].

In regards to predictors, there are numerous crime-scene variables to consider that could be beneficial in terms of characterising serial killers. For example, I propose that males rape their victims more often than females, kill with a firearm and bound their victims to a greater degree- thus considering covariates **RapedVictims**, **KilledWithGun** and **BoundVictims** seem to be candidate choices to aid in profiling killers (the meanings behind such covariates should be self explanatory and further details will be given upon application). Furthermore, victim data such as their race and gender should be useful to profile killers. Killers tend to commit murder in their own race [1] and so covariates **MaleVictims**, **FemaleVictims**, **Whitevictims**, **BlackVictims** are candidate variables to be explored.



Figure 11: U.S. Clustered By Regions With States Given

Graphic created by Sean P. O'Connor, National Geographic, May 2022.

Finally, it is worth noting that considering variation of killers at the state level may have its limitations. A large disparity is present between the geographical location in which serial killers kill, with high-killing states being California, Texas and Florida comprising of 198, 119, and 164 killers, and serial killer free state examples being New Hampshire, Wyoming and Vermont, with only one or two killers observed for each. A solution to this could be to consider killers at the regional level, instead of the state level, with state to region mappings given in figure (11).

## 8 Binary Responses

Before we apply multivariate multilevel modeling to profile serial killers, predicting their characteristics and behaviour, let us consider the univariate case in which we model a binary response variable. Our goal here is to understand modeling assumptions, estimation procedures and parameter interpretation for binary data- only normally distributed responses have been considered in previous examples. This approach is due to the fact that profiling serial killers considers discrete variables such as **Race**, **Sex**, **Motive** etc., thus understanding the univariate case for binary responses will aid us in exploring multivariate extensions. Recall the random intercept model for a continuous response variable:

$$y_{ij} = (\beta_0 + u_{0j}) + \beta_1 x_{ij} + e_{0ij},$$

where the group effects or level 2 residuals  $u_{0j}$  and the level 1 residuals  $e_{0ij}$  are assumed to be i.i.d, following normal distributions with zero means:

$$e_{0ij} \sim N(0, \sigma_{e0}^2) \text{ and } u_{0j} \sim N(0, \sigma_{u0}^2).$$

To understand the discrete case we now express the RIM model in terms of the expected value of  $y_{ij}$  for an individual in group  $j$  and with observed value  $x_{ij}$  on  $x$ :

$$E(y_{ij}|x_{ij}, u_{0j}) = (\beta_0 + u_{0j}) + \beta_1 x_{ij}. \quad (26)$$

For a binary response  $y_{ij}$ , we have  $E(y_{ij}|x_{ij}, u_{0j}) = \pi_{ij} = P(y_{ij}|x_{ij}, u_{0j} = 1)$ , where  $\pi_{ij}$  denotes the probability of the response occurring i.e., taking value “1”. In practice, binary responses are encoded as “0” or “1” for example variable **Race** could be mapped as (Black, White)  $\rightarrow$  (“0”, “1”). Ordering does not matter. A **generalised linear random intercept model** for the dependency of the response probability  $\pi_{ij}$  on  $x_{ij}$  is written as

$$F^{-1}(\pi_{ij}) = (\beta_0 + u_{0j}) + \beta_1 x_{ij}, \quad (27)$$

where  $F^{-1}$  (“F inverse”) is the link function, taken to be the inverse cumulative distribution function of a known distribution [4]. Possible link functions to choose from are “logit”, “probit”, “clog” and “log”. The advantages and disadvantages of these link functions will not be discussed here but can be found on work by Li, Jingwei 2014 on “choosing the proper link function for binary data” [10]. In the following example, we choose the popular link function “logit” also known as the log-odds link function with definition:

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right), \quad (28)$$

where  $\log$  denotes the natural logarithm and  $p$  the probability of an event occurring. Its nickname comes from the fact that this link function describes the odds on an event  $p$  from occurring or not occurring- in this application the response outcomes encoded as “1” are treated as the positive outcome  $p$  of the response. For example, a logit value of  $\text{logit}(\pi_{1ij}) = \log\left(\frac{\pi_{1ij}}{1-\pi_{1ij}}\right) = 2$  would result in a log-odds outcome of *Black:White* of  $e^2 \approx 7$  i.e., *Black* serial killers (our “1” encoded response for variable **Race**) are around 7 times more prevalent than *White* killers. In a “logit” model  $F^{-1}(\pi_{ij})$  is the log-odds that  $y = 1$  so (27) becomes

$$\log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = (\beta_0 + u_{0j}) + \beta_1 x_{ij}, \text{ with } u_{0j} \sim N(0, \sigma_{u0}^2) \text{ i.i.d copies } (u_{0j})_{j=1, \dots, m}. \quad (29)$$

## 8.1 Application

In this example we focus on illustrating model (29) applied to profiling serial killers, with the goal of understanding the mechanism behind the model, the interpretation of parameter values and assumptions used. This will result in tackling the multivariate case more intuitive and complete. Consider the following hypothesis:

- Serial killers victims, on average, comprise of victims within the same racial group as the killer. Black serial killers tend to murder black individuals and white serial killers, white individuals.

A binary response model seems like a suitable choice to answer this hypothesis. Let  $y_{ij}$  denote the race of a serial killer, given by a previously explored variable in the Radford database **Race**, a binary variable (transformed as such, with justifications given prior) taking values (*Black*, *White*) with (750, 1024) individuals and encoding (“0”, “1”) respectively. Define our predictor variable  $x_{ij}$  as **Whitevictims**, a numerical variable taking integer values between 0 and 48, with an average of 3.3 white victims per killer. 364 serial killers in the Radford database have missing entries within this predictor variable, around 20%. Consider variations in serial killer race at the state-level, grouping serial killers who kill victims with equivalent race to that of their own in each state. Thus we may write

$$\text{Race}_{ij} \sim \text{Binomial}(1, \pi_{ij}), \quad (30)$$

with a “logit” dependency of the response probability  $\pi_{ij}$  on  $x_{ij}$  written as

$$\text{logit}(\pi_{ij}) = (\beta_0 + u_{0j}) + \beta_1 \text{Whitevictims}_{ij}, \quad (31)$$

with level 2, state-level residuals  $u_{0j} \sim N(0, \sigma_{u0}^2)$  i.i.d copies  $(u_{0j})_{j=1, \dots, m}$ . In this example, our response variables follow the binomial distribution with size parameter 1. This is equivalent to the Bernoulli distribution and can be written as such; yet in the context of multilevel models, the general convention is to treat responses as binomial random variables. This is because repeated measure cases (such as students taking multiple resit papers of the same examination or patients in hospital receiving repeated blood measurements) are continuation cases to consider in this field and can be naturally extended in the binomial case by varying the sample size of the response.

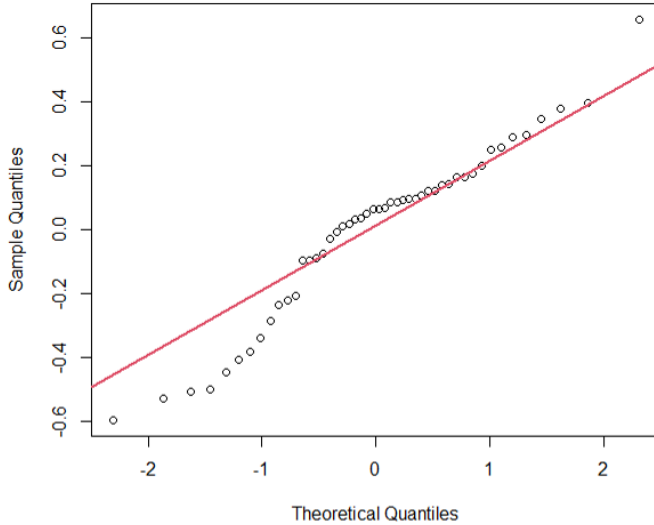


Figure 12: Binary RIM Quantile-Quantile Plot Of Level 2 Residual Error

Before we explore model findings, normality assumptions in the state-level residuals must be verified. Figure (12) shows a normal quantile-quantile plot of level 2 residual error  $\{u_{0j}\}$ . Each dot represents the random effect of each of the 50 U.S. states. The plot shows a large amount of negative skewness present from the state-level effects as evident by the small initial increase in theoretical quantiles relative to sample quantiles- dots are significantly trending below the red line. This skewness present at the state-level may be a result of small sample sized states skewing the state-level effects which will now be explored.

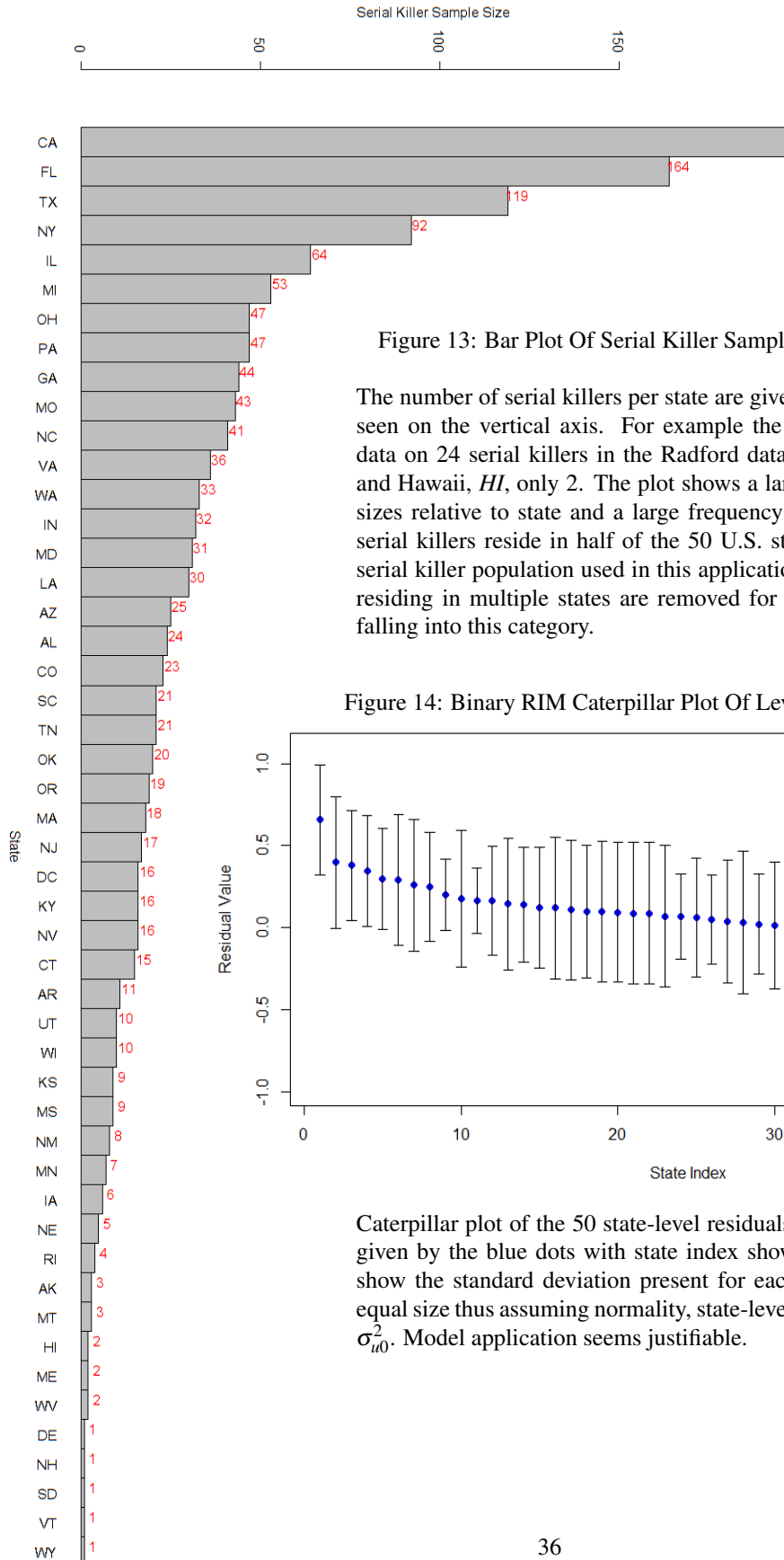
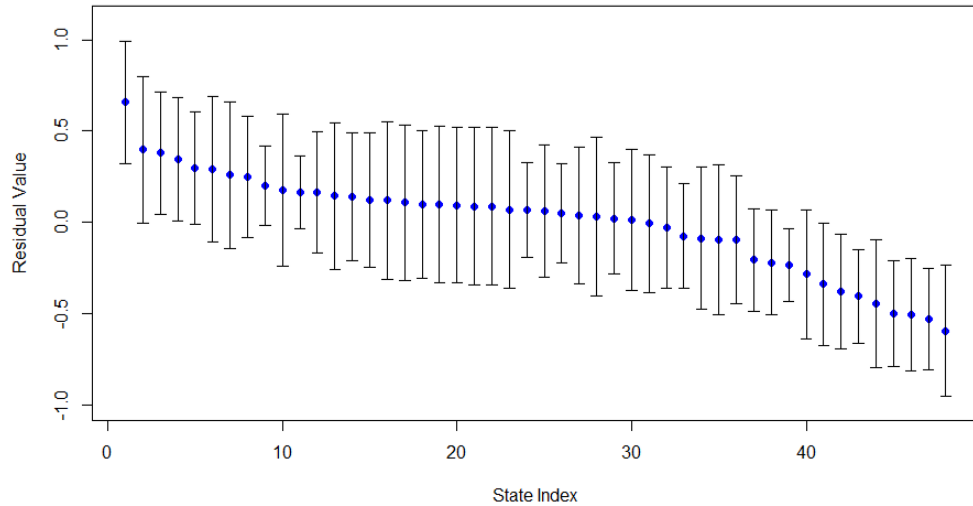


Figure 13: Bar Plot Of Serial Killer Sample Sizes By State

The number of serial killers per state are given in red, with the state initials seen on the vertical axis. For example the state Alaska, *AL*, consists of data on 24 serial killers in the Radford database, Florida, *FL*, 164 killers and Hawaii, *HI*, only 2. The plot shows a large imbalance in killer sample sizes relative to state and a large frequency of small sampled states- 254 serial killers reside in half of the 50 U.S. states, around 17% of the total serial killer population used in this application. Recall killers with victims residing in multiple states are removed for this example with 481 killers falling into this category.

Figure 14: Binary RIM Caterpillar Plot Of Level 2 Residual Error



Caterpillar plot of the 50 state-level residuals in order of descending value given by the blue dots with state index shown. The whiskers on the plot show the standard deviation present for each  $\{u_{0j}\}$ - whiskers roughly of equal size thus assuming normality, state-level residuals have equal variance  $\sigma_{u0}^2$ . Model application seems justifiable.

Race				
Predictors	Estimates	Std. Err	CI	p
(Intercept)	-0.605	0.131	-0.86– -0.35	<0.001
Whitevictims	0.371	0.038	0.30–0.45	<0.001
<b>Random Effects</b>				
$\sigma^2_{u0State}$	0.195	0.092		
Observations	1075			

Table 11: RIM Estimates For Model (31),  $\text{logit}(\text{Race}) \sim 1 + \text{Whitevictims} + (1|\text{State})$

Model parameters are estimated via IGLS (see previous section on explanation) with interpretation of fixed estimates  $\hat{\beta}_0 = -0.605$  and  $\hat{\beta}_1 = 0.371$  as follows.  $\beta_0$  is interpreted as the log-odds that  $y = 1$  when  $x = 0$  and  $u = 0$  and is referred to as the overall intercept in the linear relationship between the log-odds and  $x$ . If we take the exponential of  $\beta_0$ ,  $\exp(\hat{\beta}_0) = 0.546$ , we obtain the odds that  $y = 1$  for  $x = 0$  and  $u = 0$ . In other words, when serial killer victims are not white it seems,

on average, there is a significant difference between the killer’s race being black or white, with two-thirds of killers being of black ethnicity when no victims are white. This is an interesting finding, especially due to the white race already being prevalent in killer data- **Race** comprises of more white killers 1024:750 *white:black*.

Regarding the fixed effects of our predictor **Whitevictims**, its corresponding parameter  $\beta_1$  can be interpreted as the effect of a 1-unit change in  $x$  on the log-odds that  $y = 1$ , although it is now the effect of  $x$  after adjusting for the group effect  $u$ . If we are holding  $u$  constant, then we are looking at the effect of  $x$  for individuals within the same group so  $\beta_1$  is usually referred to as a cluster-specific effect [4]. Again,  $\exp(\hat{\beta}_1) = 1.45$  can be interpreted as an odds ratio, comparing the odds that  $y = 1$  for two individuals (in the same group) with  $x$ -values spaced 1 unit apart. In other words, we find a small yet significant trend between white killers and white victims with 60% of serial killers being of white ethnicity given a unitary increase in the number of white victims.

While  $\beta_0$  is the overall intercept in the linear relationship between the log-odds and  $x$ , the intercept for a given group  $j$  is  $\beta_0 + u_j$  which will be higher or lower than the overall intercept depending on whether  $u_j$  is greater or less than zero. The variance of the intercepts across groups is  $\text{var}(u_j) = \sigma^2_{u0}$  with estimated value  $\hat{\sigma}^2_{u0} = 0.195$  in this application. This is referred to as the between-group variance adjusted for  $x$ , the between-group residual variance, or simply the level 2 residual variance. A large variation is seen between black and white ethnic groups at the state level. In the continuous response case, we used the intraclass correlation coefficient to quantify how much level 2 variability is present in the data with respect to the total level 1, individual-level and level 2 variations seen however, there is no such statistic available in the discrete response case.

Nevertheless, serial killers victims, on average, do seem to comprise of victims within the same racial group as the killer. Black serial killers tend to murder black individuals and white serial killers white individuals, with this relationship partially being explained by the state to which the killer resides in. States comprising of mostly white serial killers tend to see an increase in the number of white victims present, with the equivalent statement holding for black serial killers. Given victim data of a crime-scene, this finding may aid in profiling the serial killer. Now that we have a deeper understanding of the mechanism behind binary univariate models, interpretation of parameter values and assumptions used, let us consider the multivariate case with the goal of determining whether multivariate multilevel modeling can be used to profile serial killers, predicting their behaviour.

## 9 Variance Component Multivariate Multilevel Model (VCMM)

Here, we consider the most simple multivariate multilevel model, the variance component multivariate multilevel model (VCMM) whereby we examine  $s$  response variables  $y_{ij}^{(0)}, \dots, y_{ij}^{(s-1)}$ , where the superscript  $\{0, \dots, s-1\}$  denotes the unique response categories and the subscripts  $i \in \{1, \dots, n\}$  and  $j \in \{1, \dots, m\}$  the  $i$ -th level 1 response of the  $j$ -th cluster- this is equivalent notation to standard multilevel models discussed in previous sections. Analogous to any multilevel framework, VCMM could be built upon in perpetuity for instance, a level 3 super-cluster could be internalised and so-forth. The variance component multivariate multilevel model considers the absence of covariates  $\beta X = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$ , for example, the variance component multivariate multilevel model with three response variables is defined as:

$$\begin{cases} y_{ij}^{(0)} &= \beta_0 + u_{0j} + e_{0ij}, \\ y_{ij}^{(1)} &= \beta_1 + u_{1j} + e_{1ij}, \text{ and} \\ y_{ij}^{(2)} &= \beta_2 + u_{2j} + e_{2ij}; \end{cases} \quad (32)$$

where the level 2, cluster level residuals given by  $u_{0j}, u_{1j}$  and  $u_{2j}$  as well as the level 1, individual level residuals given by  $e_{0j}, e_{1j}$  and  $e_{2j}$  are assumed to be multivariate normal that is,

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & & \\ \sigma_{u01} & \sigma_{u1}^2 & \\ \sigma_{u02} & \sigma_{u12} & \sigma_{u2}^2 \end{pmatrix} \right\}, \text{ and level 1 residuals} \quad (33)$$

$$\begin{pmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{e0}^2 & & \\ \sigma_{e01} & \sigma_{e1}^2 & \\ \sigma_{e02} & \sigma_{e12} & \sigma_{e2}^2 \end{pmatrix} \right\}. \quad (34)$$

This assumption is similar to standard multilevel models discussed previously however, in the multivariate case, we don't just consider the variance of random effects of responses but the covariance between responses are considered too. This is because within a certain level, responses are allowed to be dependent on other responses. Between levels however, responses are assumed to be independent observations [6]. The variance of each random effect can be read off the diagonals of the covariance matrices seen in formulas (33) and (34). For example, the variances of the level 2 random effect of response variable  $y_{ij}^{(1)}$  would be  $\sigma_{u1}^2$ , the covariance between response variables  $y_{ij}^{(1)}$  and  $y_{ij}^{(2)}$  would be given by  $\sigma_{u12}$  at the cluster level and  $\sigma_{e12}$  at the individual level. To aid us in our understanding, let us apply this model to our serial killer data in the following application.

### 9.1 Application

In this baseline case example, we ask the following question:

- How does the race, gender and motive of a serial killer vary between killers in different U.S. states? Are there relationships present not just within the race, gender and motive of a serial killer at the state level, but between the variables themselves too?

VCMM is a tool that can be used to answer this question. If the question above required us to analyse relations of race, gender and motive independently then three individual variance component models would suffice. However, we require to consider relationships between responses too thus a multivariate variance component model seems to be the preferred choice. Denote **Race**, **Sex** and **Motive** by response categories  $y^{(0)}$ ,  $y^{(1)}$  and  $y^{(2)}$  respectively, with level 1

the individual serial killer level and level 2, the **State** cluster-level. For example,  $y_{ij}^{(1)}$  denotes the sex of the  $i$ -th serial killer of the  $j$ -th cluster i.e., the  $j$ -th state. For simplicity, merge response categories as binary variables such that **Race** takes values (*Black*, *White*), **Sex** takes values (*Female*, *Male*) and **Motive** the values (*Enjoyment/power*, *Other*); with the first output having encoding “0” and the second “1”. The main justification of such variable transformations is to introduce multivariate multilevel modeling with a simplified baseline application, building upon such a baseline in further applications. The motive *Enjoyment/power* has been chosen as the baseline motive in this example due to its large prevalence with respect to other motives- 759 killers specify *Enjoyment/power* as their motive, with the second most common *Robbery*, 385, approximately half as frequent. We may now assume our response categories given to be binomial distributed random variables with unitary samples:

$$\begin{cases} \text{Race}_{ij} & \sim \text{Binomial}(1, \pi_{0ij}), \\ \text{Sex}_{ij} & \sim \text{Binomial}(1, \pi_{1ij}), \text{ and} \\ \text{Motive}_{ij} & \sim \text{Binomial}(1, \pi_{2ij}). \end{cases} \quad (35)$$

with a “logit” dependency of the response probability  $\pi_{ij}$  written as

$$\begin{cases} \text{logit}(\pi_{0ij}) & = \beta_0 + u_{0j}, \\ \text{logit}(\pi_{1ij}) & = \beta_1 + u_{1j}, \text{ and} \\ \text{logit}(\pi_{2ij}) & = \beta_2 + u_{2j}. \end{cases} \quad (36)$$

This application is analogous to that of the univariate binary response application given in the previous section, but now three responses and their relationships between each other are internalised into the model. In doing so, this adds a significant amount of complexity to model interpretation which is why this baseline case considers the exclusion of any predictors  $X$ . Before we explore our model findings, we must first check our assumptions of normality in the level 2 residuals.

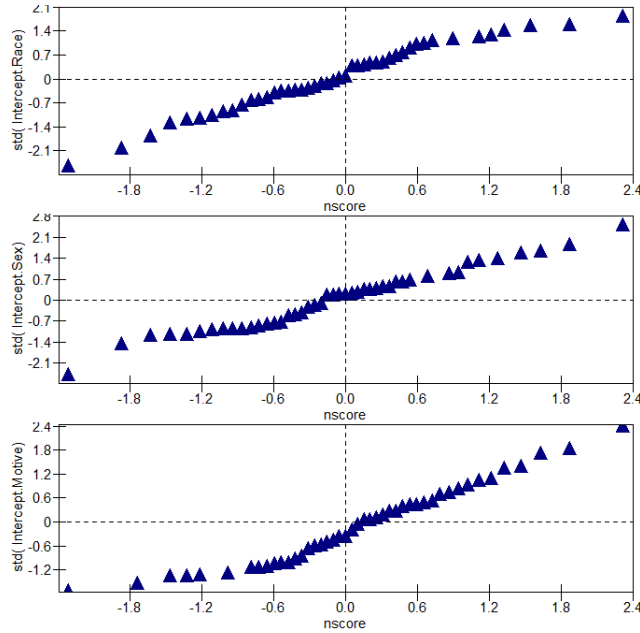


Figure 15: Binary VCMM Quantile-Quantile Plot Of Level 2 Residual Error

Figure (15) shows normal quantile-quantile plots for the level 2 error of each response with the first plot showing error at the state level for **Race**, the second **Sex** and the third plot **Motive**. Each plot consists of 50 points (the blue triangles) which are the 50 individuals state errors  $u_j$ . The horizontal axis shows a normal score (nscore) and the vertical axis the standard deviation of the residuals. The relationship between axes can be interpreted equivalently to the theoretical and sample quantiles seen in previous normal quantile-quantile plots<sup>a</sup>. There is a reasonable amount of linearity present in all response cases, justifying these state level errors seen to follow a normal distribution.

<sup>a</sup>The MLwiN software has been used directly to extract and plot the state level residuals resulting in different looking plots to that seen in prior applications whereby R was used to analyse residuals. This is because the package R2MLwiN used to apply multivariate models within R failed to extract residuals. The error ‘Error in foreign::read.dta(resfile[i]) :’ was seen in the console when attempting to extract residuals within R.

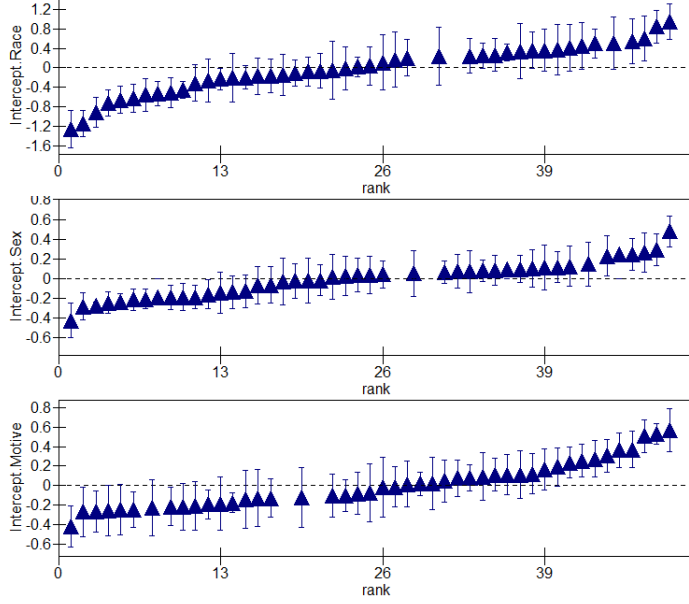


Figure 16: Binary VCMM Caterpillar Plot Of Level 2 Residual Error

Here we see caterpillar plots for the level 2 error in the equivalent order in response variables to that seen in the quantile-quantile plots above. Again, each plot consists of 50 points (the blue triangles) which are the 50 individuals state error  $u_j$ . The horizontal axis shows a residual rank and the vertical axis the raw value of the residuals. The whiskers on the plot indicate the standard deviation seen for each residual error. Within each response variable the standard deviation of the residuals are similar (to some extent). Model application seems justifiable.

Predictors	Estimates	Std. Err	CI	p
<b>Fixed Effects</b>				
(Intercept)[Race]	0.192	0.119	-0.04–0.43	0.1068
(Intercept)[Sex]	2.489	0.107	2.28–2.70	<0.001
(Intercept)[Motive]	0.339	0.078	0.19–0.49	<0.001
<b>Level 2 Random Effects</b>				
$\sigma_{u0}^2$	0.392	0.127		
$\sigma_{u1}^2$	0.060	0.078		
$\sigma_{u2}^2$	0.103	0.050		
$\sigma_{u01}$	0.126	0.079		
$\sigma_{u02}$	-0.178	0.067		
$\sigma_{u12}$	-0.009	0.046		
<b>Level 1 Random Effects</b>				
$\sigma_{e0}^2$	1.000	0.000		
$\sigma_{e1}^2$	1.000	0.000		
$\sigma_{e2}^2$	1.000	0.000		
$\sigma_{e01}$	-0.171	0.026		
$\sigma_{e02}$	-0.080	0.027		
$\sigma_{e12}$	-0.221	0.024		
Observations	1421			
Method	RIGLS			
Elapsed Time	10.11s			
Convergence	Failed			

Table 12: VCMM Estimates For Model (29),  $c(\text{logit}(\text{Race}), \text{logit}(\text{Sex}), \text{logit}(\text{Motive})) \sim 1 + (1 + \text{State})$

Estimated intercept values for the race, sex and motive of the serial killers are  $\hat{\beta}_0 = 0.192$ ,  $\hat{\beta}_1 = 2.489$  and  $\hat{\beta}_2 = 0.339$  respectively. These describe, as mentioned prior, the log-odds of the “1” response event from occurring. For example, the log-odds value for the sex of the killers can be used to find the proportion of males  $p$  that is,

$$\begin{aligned} \text{logit}(p) &= \log\left(\frac{p}{1-p}\right) \\ &= 2.489, \text{ giving} \\ p_{\text{Males}} &= 0.92. \end{aligned}$$

In other words, the intercept values help indicate how imbalanced a response variable is-  $p_{\text{Males}} = 0.92$  indicates serial killers are male dominated with around 1750 male and 150 female killers comprising the Radford database. Whilst this result is not of too much interest, this baseline application does provide some fruitful results. Table (12) displays the level 2 (state-level) random effects and level 1 (individual-level) random effects.



A statistic of note is that of the state-level variance seen regarding the race of serial killers-  $\hat{\sigma}_{u0}^2 = 0.392$  with a standard error of 0.127. As conjectured prior, a high racial disparity is seen between killers within different states. Black serial killers tend to murder victims, on average, within different states to that of white serial killers.

It is also worth noting the reasoning for unitary values seen for the level 1 variances  $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_{e2}^2 = 1$ . I suspect this is an artifact of model construction as the multivariate case seems to assume standard normality at the individual level. Serial killers are assumed to have level 1 residuals  $e \sim N(0, 1)$  in this discrete response case to produce a frame of reference and are compared accordingly, resulting in response covariance values of  $\hat{\sigma}_{e01} = -0.171$ ,  $\hat{\sigma}_{e02} = -0.080$ , and  $\hat{\sigma}_{e12} = -0.221$ . This somewhat makes sense as now these covariance values are comparable to one another for example, the race and motive of a serial killer, on average, has no significant relationship with a covariance value of  $\hat{\sigma}_{e02} = -0.080$  seen- if a killer's motive for murder is enjoyment or power then this does not provide any insight into the race of the killer in question. For the sex and motive of serial killers, there seems to be the following relation. Male killers tend to murder their victims for enjoyment more often than female killers with a fairly significant covariance between responses of  $\hat{\sigma}_{e12} = -0.221$  seen within the data. Finally, black serial killers are, on average, more likely to be male in comparison to white serial killers with a covariance of  $\hat{\sigma}_{e01} = -0.171$  seen. Since the multivariate case produced a frame of reference at the individual level then we can say that the greater the absolute value of the covariance seen between responses, the greater the effect relative to the other relations seen between responses. Whilst there are numerous examples of multivariate multilevel analysis seen in the literature, the majority cover continuous responses or a mix between discrete and continuous responses whereby this level 1 artifact of the variances taking values  $\sigma_{e0}^2 = \sigma_{e1}^2 = \sigma_{e2}^2 = 1$  is not present.

The method used to estimate model parameters of model (29) is restricted iterative generalised least squares (RIGLS) and is referenced in table (12). This is similar to iterative generalised least squares (IGLS) used to estimate model parameters for previous model applications however, RIGLS is the preferred choice for cases where large data partitions are made upon producing a hierarchical structure. In this example, clustering by **State** results in 50 level 2 cases of which, a high variation in sample size is seen- some states such as Alaska only comprise of one or two serial killer cases whilst others such as Los Angeles and California see serial killer cases in the hundreds- see figure (13). RIGLS reduces a bias that is seen in IGLS that is prevalent at the level 2 case in small sampled clusters, this also increases the convergence rate of parameter estimation. Nevertheless, application of model (29) produced estimates that did not converge with an elapsed time of 10.11 seconds. A lack of convergence would suggest that our findings produced in this application are inaccurate and unreliable however, the estimated parameter values and variances seen in table (12) are equivalent to the values seen from convergence to three significant figures. The results seen in the table are produced via the package R2MLwiN, a package within R that can perform multivariate analysis as explained in appendix (A.1)- this is where you would find a discussion of the statistical software used. The results from this package do not converge yet are identical (to three significant figures) to that of results produced via the software MLwiN directly, which do indeed converge, deeming our findings reliable. Nevertheless, this lack of convergence seen is not ideal.

A possible solution to these convergence issues is to reduce spatial data into five groups, considering clustering by U.S. regions instead of U.S. states with the state to region groupings seen in figure (11). An additional advantage of this alternative approach is the natural inclusion of multiple state killers. When clustering by state we must disregard killers that murder in multiple states due to the hierarchical linearity present in a multilevel model framework. This is not a small sample present within the dataset too- 481 out of 1902 serial killers murder in multiple states. Whilst considering a multiple-membership framework<sup>6</sup> could fix this issue [4], clustering by region enables us to still apply a multilevel structure to serial killers that kill in multiple states with killers being internalised into the region where the majority of the states they kill in reside.

---

<sup>6</sup>see appendix (A.3) for explanation of multiple-membership model structure

## 10 Random Intercepts Multivariate Multilevel Model (RIMM)

Recall the main focus of our multivariate analysis of serial killers:

- Can multivariate multilevel modeling be used to profile serial killers, predicting their characteristics and behaviour?

From understanding a baseline model that considered no predictor variables (VCMM), we will now try to profile serial killers using crime-scene data. The **random intercepts multivariate model (RIMM)** can be used to answer our question above. This model is equivalent to that of the variance multivariate model with the addition of covariates  $\beta X = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$  where  $\beta_0, \dots, \beta_p \in \mathbb{R}$  are input parameter constants. For example, the random intercepts multivariate model with three response variables is defined as:

$$\begin{cases} y_{ij}^{(0)} &= (\beta_0 + u_{0j}) + \sum_{l=1}^p \beta_l x_{lij} + e_{0ij}, \\ y_{ij}^{(1)} &= (\beta_1 + u_{1j}) + \sum_{l=1}^p \beta_l x_{lij} + e_{1ij}, \text{ and} \\ y_{ij}^{(2)} &= (\beta_2 + u_{2j}) + \sum_{l=1}^p \beta_l x_{lij} + e_{2ij}; \end{cases} \quad (37)$$

where  $\beta_l$  is the constant fixed effect of the  $l$ -th covariate  $l \in 1, \dots, p$  and where the level 2, cluster level residuals given by  $u_{0j}, u_{1j}$  and  $u_{2j}$  as well as the level 1, individual level residuals given by  $e_{0j}, e_{1j}$  and  $e_{2j}$  are assumed to be multivariate normal that is,

$$\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u0}^2 & & \\ \sigma_{u01} & \sigma_{u1}^2 & \\ \sigma_{u02} & \sigma_{u12} & \sigma_{u2}^2 \end{pmatrix} \right\}, \text{ and level 1 residuals} \quad (38)$$

$$\begin{pmatrix} e_{0ij} \\ e_{1ij} \\ e_{2ij} \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{e0}^2 & & \\ \sigma_{e01} & \sigma_{e1}^2 & \\ \sigma_{e02} & \sigma_{e12} & \sigma_{e2}^2 \end{pmatrix} \right\}. \quad (39)$$

The explanation of VCMM notation in section (9) is analogous to the random intercept multivariate model- I refer you to this section for information on parameter interpretation.

### 10.1 Application

The crime-scene data of interest in this application are the covariates **RapedVictims**, **KilledWithGun** and **BoundVictims**. The meanings behind all three variables are self explanatory. These variables are encoded with output (“0”, “1”) where “1” denotes the serial killer in question did indeed rape their victims, killed with a gun or bound their victims and “0” the complement outcome. Table (13) shows the number of outcomes for all three covariates, as well as the amount of missing values seen.

Variable	No	Yes	Missing Values
<b>RapedVictims</b>	1053	703	146
<b>KilledWithGun</b>	842	425	635
<b>BoundVictims</b>	1164	312	426

Table 13: Data Matrix Of Crime-scene Data

Here, we consider the same responses (**Race**, **Sex** and **Motive**) to that used in our baseline application seen in the previous section. I refer you to this previous application for more information regarding these responses. In short,

similar to the VCMM application, we may assume our response categories given to be binomial distributed random variables with unitary samples:

$$\begin{cases} \text{Race}_{ij} & \sim \text{Binomial}(1, \pi_{0ij}), \\ \text{Sex}_{ij} & \sim \text{Binomial}(1, \pi_{1ij}), \text{ and} \\ \text{Motive}_{ij} & \sim \text{Binomial}(1, \pi_{2ij}). \end{cases} \quad (40)$$

However, we are now profiling killers using crime-scene data, this gives a “logit” dependency of the response probability  $\pi_{ij}$  written as

$$\begin{cases} \text{logit}(\pi_{0ij}) & = (\beta_0 + u_{0j}) + \beta_3 \text{RapedVictims}_{ij} + \beta_4 \text{KilledWithGun}_{ij} + \beta_5 \text{BoundVictims}_{ij}, \\ \text{logit}(\pi_{1ij}) & = (\beta_1 + u_{1j}) + \beta_6 \text{RapedVictims}_{ij} + \beta_7 \text{KilledWithGun}_{ij} + \beta_8 \text{BoundVictims}_{ij}, \text{ and} \\ \text{logit}(\pi_{2ij}) & = (\beta_2 + u_{2j}) + \beta_9 \text{RapedVictims}_{ij} + \beta_{10} \text{KilledWithGun}_{ij} + \beta_{11} \text{BoundVictims}_{ij}. \end{cases} \quad (41)$$

An explanation of model (41) is as follows. Predictor variable  $\text{KilledWithGun}_{ij}$  is the outcome of the  $i$ -th serial killer in the  $j$ -th cluster on whether this specific serial killer murdered his victims with a gun or not, with “1” indicating the killer used a gun and “0” he did not. The same interpretation can be given to variables  $\text{RapedVictims}_{ij}$  and  $\text{BoundVictims}_{ij}$ . Again, since we are interested in binary responses **Race**, **Sex** and **Motive**, a “logit” link function is used- this is the same link function used in the baseline model.

As with all multivariate multilevel analysis seen, in this application we group killers into the 50 U.S. states whereby the residuals of killers at the state-level are given by  $u_{0j}$ ,  $u_{1j}$  and  $u_{2j}$  for a killer’s race, sex and motive respectively. Before model application, these residuals are checked for presence of normality as usual practice dictates, with normal quantile-quantile plots and caterpillar plots of the state-level residuals seen below. I refer you to figures (15) and (16) for explanations of these plots- in short, model application seems justifiable.

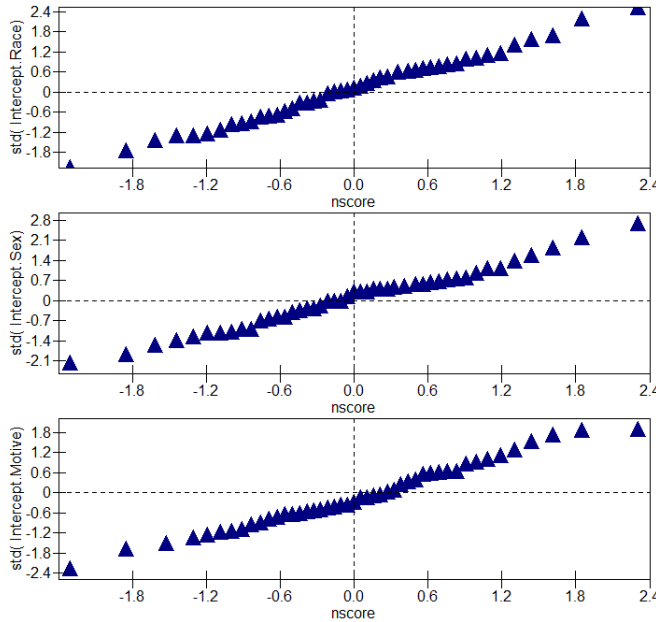


Figure 17: Binary RIMM Quantile-Quantile Plot Of Level 2 Residual Error

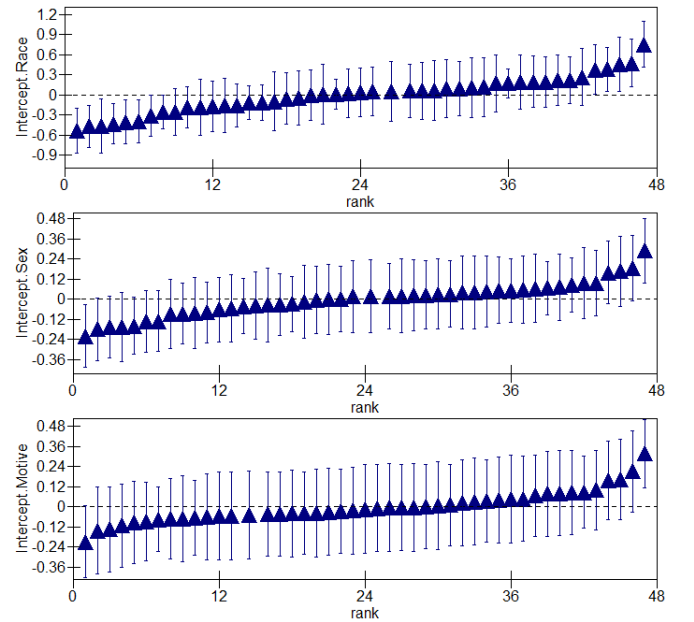


Figure 18: Binary RIMM Caterpillar Plot Of Level 2 Residual Error

Predictors		Estimates	Std. Err	CI	p		
Fixed Effects							
(Intercept)[Race]		0.749	0.158	0.44–1.06	<0.001		
(Intercept)[Sex]		0.895	0.153	0.59–1.20	<0.001		
(Intercept)[Motive]		1.375	0.159	1.06–1.69	<0.001		
RapedVictims(Yes)[Race]		-0.766	0.179	-1.12– -0.42	<0.001		
RapedVictims(Yes)[Sex]		4.367	0.965	2.47–6.26	<0.001		
RapedVictims(Yes)[Motive]		-3.008	0.206	-3.41– -2.60	<0.001		
KilledWithGun(Yes)[Race]		-0.839	0.176	-1.18– -0.49	<0.001		
KilledWithGun(Yes)[Sex]		1.380	0.277	0.84–1.92	<0.001		
KilledWithGun(Yes)[Motive]		0.782	0.220	0.35–1.21	<0.001		
BoundVictims(Yes)[Race]		0.783	0.207	0.38–1.19	<0.001		
BoundVictims(Yes)[Sex]		0.995	0.471	0.07–1.92	0.0346		
BoundVictims(Yes)[Motive]		-0.596	0.245	-1.08– -0.12	0.0151		
Level 2 Random Effects		Estimates	Std. Err	Level 1 Random Effects		Estimates	Std. Err
$\sigma_{u0}^2$		0.207	0.108	$\sigma_{e0}^2$		1.000	0.000
$\sigma_{u1}^2$		0.048	0.109	$\sigma_{e1}^2$		1.000	0.000
$\sigma_{u2}^2$		0.070	0.089	$\sigma_{e2}^2$		1.000	0.000
$\sigma_{u01}$		0.070	0.081	$\sigma_{e01}$		-0.087	0.037
$\sigma_{u02}$		-0.023	0.072	$\sigma_{e02}$		-0.074	0.037
$\sigma_{u12}$		-0.003	0.070	$\sigma_{e12}$		-0.076	0.036
		Observations		785			
		Method		IGLS			
		Elapsed Time		3.12s			
		Convergence		Successful			

Table 14: RIMM Estimates For Model (41),  $c(\text{logit}(\text{Race}), \text{logit}(\text{Sex}), \text{logit}(\text{Motive})) \sim 1 + \text{RapedVictims} + \text{KilledWithGun} + \text{BoundVictims} + (1 \mid \text{State})$

The model produces estimates for the parameters seen in (41) by IGLS, not RIGLS as seen in the baseline application. This is because convergence issues were seen (and explained) in the baseline case which prompted the use of RIGLS due to the benefits seen in convergence from this parameter estimation method. Fortunately, model (41) experiences no convergence issues with an elapsed time of convergence with IGLS of 3.12s as seen in table (14). This table also show estimates of all parameters in the model with the reference of each estimate to their corresponding  $\beta$  being self explanatory. For example, the fixed effect *KilledWithGun[Race]* corresponds to the parameter  $\hat{\beta}_4 = -0.839$ . Similar to the baseline application, these parameter estimates can be interpreted as the log-odds of that specific event occurring i.e. when the log-odds of the when response  $y = 1$  and the covariate  $x = 1$ . Following our example on the relation between a killer's race and whether they killed with a gun, the parameter  $\hat{\beta}_4 = -0.839$  suggests, on average, the majority of serial killers that use a gun to kill their victims are black relative to the amount of white serial killers who use a gun. To show this clearly it is useful to construct the following table of outcomes.

Variable	“0” Encoded	“1” Encoded
<b>Race</b>	<i>Black</i>	<i>White</i>
<b>Sex</b>	<i>Female</i>	<i>Male</i>
<b>Motive</b>	<i>Enjoyment/power</i>	<i>Other</i>
<b>RapedVictims</b>	<i>No</i>	<i>Yes</i>
<b>KilledWithGun</b>	<i>No</i>	<i>Yes</i>
<b>BoundVictims</b>	<i>No</i>	<i>Yes</i>

Table 15: Variable Outcome Encoding

A log-odds estimate of  $\hat{\beta}_4 = -0.839$  gives the following probability  $p$  that is, the probability of the response  $y = 1$  when  $x = 1$ . We have

$$\begin{aligned} \text{logit}(p) &= \log\left(\frac{p}{1-p}\right) \\ &= -0.839, \text{ giving} \\ p_{y=1|x=1} &= 0.302. \end{aligned}$$

Using the table of outcomes, table (15), our findings seem to suggest that on average, the majority of serial killers that use a gun to kill their victims are black relative to the amount of white serial killers who use a gun.

This result is significant too with a confidence interval  $CI = [-1.18, -0.49]$ , spanning a negative domain only, and a p-value  $p < 0.001 \ll 0.05$ , very small. In actuality, all parameter estimates for model (41) seem to be statistically significant, with confidence intervals spanning values reasonably far from zero and small p values present - this could be questioned for the predictor variable **BoundVictims** in regards to the responses **Sex** and **Motive**.

There are many interesting findings that could be used to profile serial killers seen in this model application. The three most dominant findings (the scenarios with the largest log-odds absolute values) are *RapedVictims[Sex]*, *RapedVictims[Motive]* and *KilledWithGun[Sex]* with log-odds estimated values of 4.367, -3.008 and 1.380 respectively. We can apply the same procedure that has been explained in detail with the scenario *KilledWithGun[Race]* to infer the following characteristics of serial killers.

The vast majority of victims that are seen at a serial killer’s crime-scene of which, have indications of rape, seem to have been killed by a male serial killer (98.7% of the times this scenario occurred, the serial killer is a male). This observation is not too surprising, since serial killers in general are male dominated with 1745 male serial killers and 150 female killers comprising the Radford database- 92% of the total sample seen in the database are males. Furthermore, without wanting to delve into the morbid practicalities of serial killer crime-scenes, one would suspect it to be self evident given a raped victim body to determine the gender of the killer whom did such an act- trying to predict the gender of the killer for this scenario using this quantitative finding seems like the wrong approach.

A somewhat more useful finding is that serial killers, on average, rape their victims for reasons of enjoyment and power the vast majority of the time, with a log-odds of this scenario occurring seen to be -3.008- roughly 95% of the time. Determining the motive behind a killer is vital information that could be used by criminal investigation units to aid in the capturing of the serial killer. Whilst predicting a serial killer to rape their victims out of enjoyment and power is not too surprising, the high proportion of killers with this motive seen is. One would expect rape out of motives such as *Anger* or *Revenge* to be more prevalent that what our findings suggest.

Finally, the parameter *KilledWithGun[Sex]* with a log-odds of 1.380 suggests that on average, male serial killers use a firearm more often than female killers with a probability of a killer being of the male gender given the use of a gun being roughly 80%. However, as discussed previously the vast majority of serial killers are male to begin with, 92%, this finding should be interpreted with care. In short, when determining the gender of a serial killer it seems that we are always inclined to suspect the killer to be male and so using crime-scene data to try to predict the gender of a killer is not an exercise that is all too useful.

As seen in the baseline application prior in regards to the variability between serial killers within different states, the race of a serial killer seems to vary at the state-level with a variance of  $\hat{\sigma}_{u0}^2 = 0.207$  and a standard error of 0.108. Again, a high racial disparity is seen between killers within different states. Black serial killers tend to murder victims, on average, within different states to that of white serial killers. It is worth noting this variation has lessened with the inclusion of crime-scene data- a variance of  $\hat{\sigma}_{u0\text{Baseline}}^2 = 0.392$  is seen in the baseline case compare to  $\hat{\sigma}_{u0\text{Crime-scene}}^2 = 0.207$  for this crime-scene application. This may be due to this crime-scene application using almost half as many observations to that seen in the baseline example- 785 observations compared to 1421. The addition of predictor variables results in the exclusion of data entries with missing values, especially due to the use of the predictor **KilledWithGun** that sees 635 missing values (see table (13) for more information on missing values).

## 11 Conclusion

### 11.1 Multilevel Modeling

Given a dataset containing 1902 serial killers spanning 80 variables, we have **demonstrated various multilevel models and how they might be applied in criminology**. Variance component, random intercept and random slope models were gradually introduced, and have been shown to be natural extensions of the generalized linear model framework. Applying such models to serial killer data has given the following findings:

1. While most variability in the age at which a serial killer commits their first murder stems from differences between killers, a substantial amount of variability is explained by the motive behind such killings. Killers relating to *Organised crime* and *Cult* related serial killings seem to start at earlier ages than average (27 years). Vice versa, *Financial/personal gain* incentives and *Black widow* killings occur at later ages than average (section (3)).
2. Black ethnic groups tend to start killings at an earlier age compared to white serial killers, with the data suggesting around two and a half years difference between the two groups (section (4) and (5)).

Not only is multilevel modeling a justified approach to aid in understanding the criminal mind, but this model framework seems to produce more fruitful results (in specific contexts) in comparison to standard statistical modeling practices such as applying generalized linear models. The reasoning behind the use of multilevel models is best argued in general. There are two main criteria that should be considered prior to application. Firstly, are your response variables independent from one another? One of the assumptions behind generalized linear models is independence between responses  $\{y_i\}$ . If this independence assumption is violated, then applying generalized linear models could lead to unreliable findings[6]. Multilevel models are a tool that could fix this dependency issue by partitioning the data into natural clusters such that within each cluster, the responses are indeed independent from each other. This is best illustrated with an example which is discussed in appendix (A.2). To determine how to partition the data is a reasonable question in its own right- usually the context of the data should suggest a natural way to group the data for instance, grouping students by school, patients by hospital etc. The second criteria that should be considered prior to applying multilevel modeling is: what finding do we wish to explore in our data analysis? If we are considering variations seen between responses then multilevel modeling seems to be an elegant approach. For example, recall the simplicity of the random slope model applied to exploring variations between the race of a serial killer and the motive behind their killings. It allowed us to get a measure of how much the effect of race varies between motives by estimating a single parameter (the level 2 variance of  $u_{0j}$ ). As seen, a similar analysis with a fixed effect model has required us to fit a model with many more parameters (an interaction effect for race with 17 different motives), leading to a much more complex and less interpretable model. If the question is simply "Does the effect of [response] vary by [group] and, if so, by how much?" then the multilevel approach seems quite attractive.

### 11.2 Multivariate Multilevel Modeling

Multivariate multilevel modeling has been subsequently explored with the purpose of determining whether **variations seen between serial killers can be used to profile killers, predicting their characteristics and behaviour**. Crime-scene data has been used with the goal of predicting a killer's race, gender and motive, together with gaining insight into how these characteristics vary within different U.S. states. The findings of such data analysis is as follows:

1. Serial killers victims, on average, do seem to comprise of victims within the same racial group as the killer. Black serial killers tend to murder black individuals and white serial killers white individuals, with this relationship partially being explained by the state to which the killer resides in. States comprising of mostly white serial killers tend to see an increase in the number of white victims present, with the equivalent statement holding for black serial killers (section (9) and (10)).
2. On average, the majority of serial killers that use a gun to kill their victims are black relative to the amount of white serial killers who use a gun (section (10)).

3. The vast majority of victims that are seen at a serial killer's crime-scene of which, have indications of rape, seem to have been killed by a male serial killer (98.7% of the times this scenario occurred, the serial killer is a male, see section (10)).
4. Serial killers, on average, rape their victims for reasons of enjoyment and power the vast majority of the time (95% of the time), with motives such as *Anger* or *Revenge* not being as prevalent as one would expect (section (10)).
5. Male serial killers use a firearm more often than female killers (80% of the time, see section (10)).

Variations seen between serial killers do indeed seem to be useful in aiding in the profiling of killers, predicting their characteristics and behaviour. This is especially true for the race of a serial killer which has been seen to vary significantly throughout the states a serial killer resides in- state-level random effects for race were  $\hat{\sigma}_{u0VCMM}^2 = 0.392$  and  $\hat{\sigma}_{u0RIMM}^2 = 0.207$  for variance component and random intercept multivariate analysis applied in sections (9) and (10). In regards to the sex and motive of a killer, it seems no meaningful variation is seen at the state-level- our state-level random effects were small with variance estimates of  $\hat{\sigma}_{u1}^2 = 0.048$  and  $\hat{\sigma}_{u2}^2 = 0.070$  for the gender and motive of serial killers respectively. This is a useful observation in its own right since now we may deduce standard multivariate analysis may suffice in regards to treating the gender and motive of serial killers as a response. Finally, some of our findings listed above could be questioned. For instance whilst it does indeed seem true that male serial killers use a firearm more often than female serial killers, this statement is slightly misleading. The vast majority of serial killers seen in the Radford database are men, more than 90%. Given this dominant male presence, the finding that 80% of the time a firearm is used the killer is deemed to be male could be argued to infer that females use a firearm more often than males relative to population size. If they were equally likely then 90% of the time (not 80%) a male would have been seen to use a firearm. In short, when applying statistics to explore and analyse data, it is important to be precise when discussing any model findings and militant when checking modeling assumptions; for as Mark Twain, a famous American writer once said, "facts are stubborn things, but statistics are pliable".

### 11.3 Further Research

A natural continuation to our multilevel modeling journey is to consider the random slopes multivariate multilevel model case. This multivariate model allows the possibility that the effect of a covariate on the response might also vary between clusters. Findings from a random slopes model could be compared to its random intercept counterpart that has been analysed and discussed in section (10). An additional path to consider is to explore standard multivariate analysis applied to criminology, comparing such findings to that which have been produced in this research. For instance, no meaningful variation was seen between the gender and motive of serial killers and the state to which the serial killer killed their victim. Acquiring some standard multivariate analysis as a baseline to compare to the multilevel multivariate case would be a useful exercise, gaining insight into the justification of a multilevel structure applied to serial killers.

Furthermore, regarding the lack of convergence seen in the results of the VCMM application in section (9.1), a solution that could be explored is reducing spatial data into five groups, considering clustering by U.S. regions instead of U.S. states. An additional advantage of this alternative approach is the natural inclusion of multiple state killers. When clustering by state we must disregard killers that murder in multiple states due to the hierarchical linearity present in a multilevel model framework. This is not a small sample present within the dataset too- 481 out of 1902 serial killers murder in multiple states. Clustering by region enables us to still apply a multilevel structure to serial killers that kill in multiple states with killers being internalised into the region where the majority of the states they kill in reside.

It is also worth noting some technicalities that would be improved upon throughout our application of multilevel models. In regards to analysis seen whereby we grouped killers by their motives, the serial killers with *unknown* motives should have been removed from the data and treated as missing values. For a serial killer to have an unknown motive, it does not necessarily mean they possess a motive different to the other 17 motives present. Additionally, all analysis that considered the race of the serial killer categorized them as *White* or *Black*- an additional 113 data entries could have been used that is, serial killers with *Hispanic Asian* or *Native American* ethnicity if instead we categorized killers as *White* or *Non-White*. This alternative approach may have led to more reliable findings.

## References

- [1] Michael Aamodt, Terence Leary, and Larry Southard. Radford/fgcu annual report on serial killer statistics: 2020, 06 2020.
- [2] Nicole H. Augustin, Erik-Andre Sauleau, and Simon N. Wood. On quantile quantile plots for generalized linear models. *Computational Statistics and Data Analysis*, 56(8), 2012.
- [3] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*, 2014.
- [4] University Bristol. *Multilevel Modelling online course: LEMMA VLE Centre for Multilevel Modelling*. ([www.cmm.bris.ac.uk/lemma/](http://www.cmm.bris.ac.uk/lemma/)), 2012.
- [5] Julian J Faraway. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC, 2016.
- [6] Harvey Goldstein. *Multilevel statistical models*. John Wiley & Sons, 2011.
- [7] Ellen L Hamaker, Pascal van Hattum, Rebecca M Kuiper, and Herbert Hoijtink. Model selection based on information criteria in multilevel modeling. *Handbook of advanced multilevel analysis*, pages 231–255, 2011.
- [8] Joop J Hox, Mirjam Moerbeek, and Rens Van de Schoot. *Multilevel analysis: Techniques and applications*. Routledge, 2017.
- [9] Don Jacobs. *Analyzing Criminal Minds: Forensic Investigative Science for the 21st Century: Forensic Investigative Science for the 21st Century*. ABC-CLIO, 2011.
- [10] Jingwei Li et al. *Choosing the proper link function for binary data*. PhD thesis, 2014.
- [11] Fernando Martinez. Using machine learning to analyze patterns in serial killers. 2019.
- [12] Shinichi Nakagawa, Paul CD Johnson, and Holger Schielzeth. The coefficient of determination  $r^2$  and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134):20170213, 2017.
- [13] Michael Omi and Howard Winant. Thinking through race and racism, 2009.
- [14] James L. Peugh. A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1):85–112, 2010.
- [15] Jose C. Pinheiro and Douglas M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, 4(1):12–35, 1995.
- [16] Hugo Quené and Huub Van den Bergh. On multi-level modeling of data from repeated measures designs: A tutorial. *Speech communication*, 43(1-2):103–121, 2004.
- [17] Jon Rasbash, Chris Charlton, Kelyvn Jones, and Rebecca Pillinger. Manual supplement for mlwin version 3.01. 2017.
- [18] Nigel Rice and Alastair Leyland. Multilevel models: applications to health data. *Journal of health services research & policy*, 1(3):154–164, 1996.
- [19] Mark J. Schervish. P values: What they are and what they are not. *The American Statistician*, 50(3):203–206, 1996.
- [20] Holger Schielzeth. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, 1(2):103–113, 2010.
- [21] Samaradasa Weerahandi. Generalized confidence intervals. *Journal of the American Statistical Association*, 88(423):899–905, 1993.



## 12 Appendix A

### 12.1 (A.1) Statistical Software Used

The programming language R has been used exclusively throughout model applications discussed in this work, with the code commented and given in equivalent order of model application seen in appendix (B). Various packages have been used within R and can be seen at the start of the code. For multilevel modeling, the lme4 package was used primarily, for the use of the lmer function within the package. This function enables us to apply multilevel models within R. Explanation of lmer notation used throughout my discussion is seen below. For the multivariate multilevel analysis seen in the closing sections of our discussion, the software MLwiN was used. This is due to R not supporting multivariate multilevel modeling analysis. Fortunately, an R package R2MLwiN is available which contains the function runMLwiN, this enables a command interface that may execute the MLwiN software within an RStudio environment. In doing so, the annotated code in appendix (B) shows all analysis, both for the multilevel and multivariate cases. To run the code successfully the MLwiN software must be downloaded. MLwiN is free to download for UK-academics only using the following link: <http://www.bristol.ac.uk/cmm/software/mlwin/download/>.

#### 12.1.1 Linear Mixed-Effects Models (lmer) Formula Notation

A two-sided linear formula object describing both the fixed effects and random effects part of the model, with the response on the left of a ~ operator and the terms, separated by + operators, on the right. Random-effects terms are distinguished by vertical bars (|) separating expressions for design matrices from grouping factors[3]. In other words the general formulation approach can be read as follows:

$$\text{response} \sim 1 + \text{fixed effects} + (1 + \text{random effects} | \text{cluster variable}).$$

An example using applications provided in this work is as follows. The lmer formula **AgeFirstKill** ~ 1 describe a standard regression model with no predictor variable present. The formula **AgeFirstKill** ~ 1 + (1|**Motive**) shows a variance components model whereby we group serial killers by motive. **AgeFirstKill** ~ 1 + **Race** + (1|**Motive**) a random intercept model and **AgeFirstKill** ~ 1 + **Race** + (1 + **Race**|**Motive**) a random slopes model. Random effects within our covariate **Race** are seen in the random slopes model and are not seen in the random intercept model.

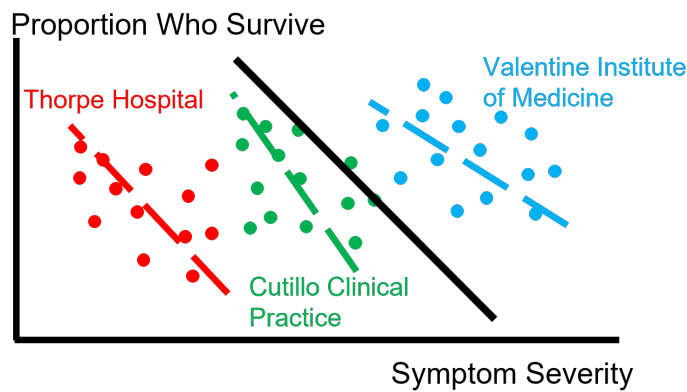
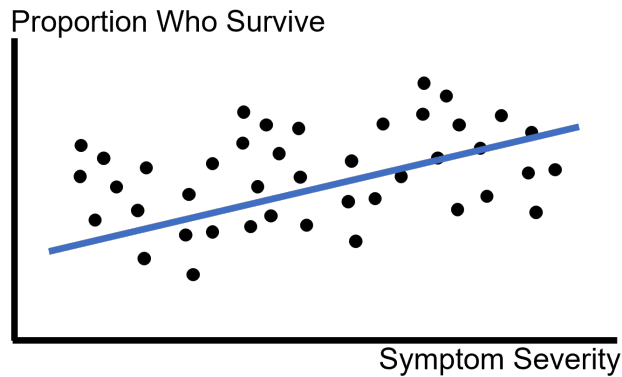
#### 12.1.2 Multivariate Multilevel Models (runMLwiN) Formula Notation

The multivariate formulation uses the equivalent operations to that of the linear mixed-effects models (lmer) and so I refer you to section (12.1.1) for details regarding operators. Where the differences arise are in the responses. In standard multilevel modeling we were only interested in a single response but in multivariate analysis multiple responses are used. These are given in a vector form with the link function specified for each of the responses. For example, a model consisting of three responses would have the general formulation:

$$c(\text{link}(\text{response}_1), \text{link}(\text{response}_2), \text{link}(\text{response}_3)) \sim 1 + \text{fixed effects} + (1 + \text{random effects} | \text{cluster variable}).$$

For example,  $c(\text{logit}(\text{Race}), \text{logit}(\text{Sex}), \text{logit}(\text{Motive})) \sim 1 + (1 | \text{State})$  is a binary VCMM for responses **Race**, **Sex** and **Motive** whereby serial killers are grouped by **State** and the responses have a “logit” link function applied.

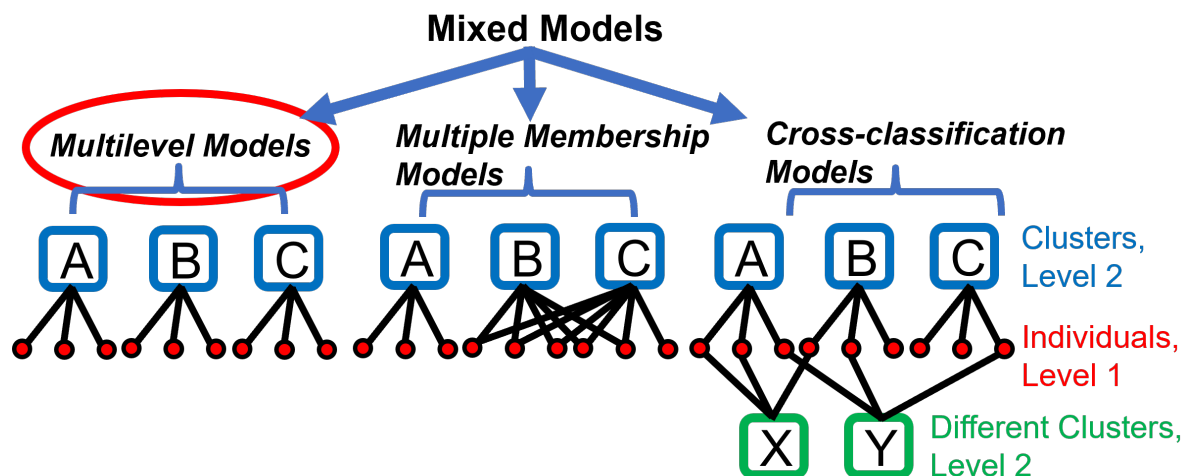
## 12.2 (A.2) RIM and RSM Illustrative Example



The two plots show an illustrative example of the random slopes model applied to the context of healthcare. The dots represent individual patients who are diseased with their symptom severity on the horizontal axis and survival rate displayed on the vertical axis. Applying standard regression as seen by the blue line in the first plot suggests that the more diseased a patient is the greater chance that patient has of surviving his/her illness.

This goes against our common sense understanding of health- you would expect the greater the symptom severity the lower the survival rate of a patient. Standard regression, in this artificial yet motivating example provides unreliable results due to the natural clusters formed within patient data. Patients are committed to three different hospitals, the red, green and blue hospitals seen in the second plot. These three hospitals commit patients with varying degrees of symptoms severity- red low symptoms, green moderate symptoms and blue severe. Internalizing these natural clusters seen within the patient data results in more appropriate findings as when we consider a specific hospital we do indeed see a decrease in survival rate as patient symptoms increase. This example shows a random slopes model as the effect of the covariate, symptom severity, is allowed to vary between hospitals- the slopes of the red, green and blue hospital are allowed to be different. In a random intercept model, these slopes are assumed to be fixed.

### 12.3 (A.3) Model Structure Visualisation



Multilevel models are a member of the mixed models family, whereby individuals seen at level 1 are nested within a level 2 cluster. This relation is shown in the graphic above by the red dots, the individuals, being linked to a specific group- either cluster A, B or C as shown by the black line. For more complex model structures, multiple membership models and cross-classification models could be considered with their model structures given in the graphic. As the name suggests, multiple membership models allow for individuals to reside in multiple clusters. In the context of serial killers, the relationship of serial killers who murder their victims in multiple states could be considered. Cross-classification models allow for the inclusion of more than one cluster variable for instance, throughout our discussion of serial killers, we considered cases in which we grouped killers by motive or grouped killers by state- cross-classification enables us to group killers by motive and state simultaneously.

## 13 Appendix B

```
#####  
# Step 1: Import the data-----  
#####  
# 1.1)-----  
# clears all data and plots from your current Global Environment  
rm(list=ls())  
cat("\014")  
  
# packages  
library('ggplot2')  
library('tidyverse')  
library('fitdistrplus')  
library('lme4')  
library('flexplot')  
library('zoom')  
library('sjPlot')  
library('predictmeans')  
library('BayesFactor')  
library('fastDummies')  
library('stringr')  
library('R2MLwiN')  
library('nnet')  
  
# working directory  
setwd("/Users/marcu/OneDrive/Desktop/Uni/Dissertation/code")  
  
# read-in data  
df <- get(load('killersfull(Marcus).Rdata'))  
rm(killersfull)  
get(load('dates.Rdata'))  
df <- cbind(df,dates)  
rm(dates)  
#####  
# Step 2: Exploration -----  
#####  
# 2.1) Imputation -----  
  
df$YearOfDeath[500] <- median(df$YearOfDeath)  
df$ConfessedKills[1758] <- median(df$ConfessedKills)  
  
df$Race <-ifelse(df$Race=='White','White',  
                ifelse(df$Race=='Black','Black',NA))  
df$Race <- as.factor(df$Race)  
  
df$Motive <- as.factor(df$Motive)
```

```

#####
# step 3: Modeling -----
#####

# 3.1) Linear regression -----
# Y = AgeFirstKill, X = Sex

# checking assumptions
plot(fitdist(as.vector(na.omit(df$AgeFirstKill)), "norm",
             discrete=F), breaks=50)

# model 1
AgeFirstKill_model_1 <- lm(AgeFirstKill ~ 1, data = df)

# model 2
AgeFirstKill_model_2 <- lm(AgeFirstKill ~ Sex,
                           data = df)

# summary
summary(AgeFirstKill_model_1)
tab_model(AgeFirstKill_model_1)
summary(AgeFirstKill_model_2)
tab_model(AgeFirstKill_model_2)

# residual assumption checking
par(mfrow = c(1,1))
par(mar = c(4, 4, 4, 4))
plot(fitdist(AgeFirstKill_model_2$residuals, "norm",
             discrete=F), breaks=30)

# Independence assumption checking, cluster=Motive
# plot
par(mfrow = c(1,1))
par(mar = c(4, 4, 4, 20))
plot(as.numeric(df$Motive)~df$AgeFirstKill, xlab='AgeFirstKill', ylab='Motive',
     pch=19, col= df$Motive, yaxt='n')
axis(4, at = 1:18, labels = paste(1:18, levels(df$Motive)), las = 2)
abline(v = 27.64, lty = 2)
axis(3, at =27.64, labels = expression(paste(widehat(beta)[0], " = 27.64")))

# 3.2) Variance components model -----
# Y=AgeFirstKill Cluster=Motive

# model 3
AgeFirstKill_model_3 <- lmer(AgeFirstKill ~ 1 + (1 | Motive),
                             data = df, REML=F)
summary(AgeFirstKill_model_3)
tab_model(AgeFirstKill_model_3)

# checking assumptions
randomeffects <- ranef(AgeFirstKill_model_3)
u <- randomeffects[["Motive"]][["(Intercept)"]]

```

```

par(mfrow = c(1,2))
par(mar = c(4, 4, 4, 4))
qqnorm(resid(AgeFirstKill_model_3))
qqline(resid(AgeFirstKill_model_3), col = 2,lwd=2,lty=1)
qqnorm(u)
qqline(u, col = 2,lwd=2,lty=1)

# plot 2
par(mfrow = c(1,1))
par(mar = c(4, 4, 4, 20))
plot(as.numeric(df$Motive)~df$AgeFirstKill,xlab='AgeFirstKill',ylab='Motive',
     pch=19,col= df$Motive,yaxt='n')
axis(4, at = 1:18, labels = paste(1:18,levels(df$Motive)), las = 2)
abline(v = 27.87, lty = 2)
axis(3, at =27.87, labels = expression(paste(widehat(beta)[0], " = 27.87")))
points(27.87 + u, 1:18, pch = "|", cex=2)
# zoomed in plot
#zm()

# 3.3) Random Intercepts model-----
rm(list=setdiff(ls(), "df"))
dev.off()

# model
baseline = lmer(AgeFirstKill~1 +(1|Motive), data=df)
icc(baseline)

fixed_slopes = lmer(AgeFirstKill~Sex +(1|Motive), data=df,REML=F)
icc(fixed_slopes)
summary(fixed_slopes)
tab_model(fixed_slopes)

# assumption checking
randomeffects <- ranef(fixed_slopes)
u <- randomeffects[["Motive"]][["(Intercept)"]]

par(mfrow = c(1,2))
par(mar = c(4, 4, 4, 4))
qqnorm(resid(fixed_slopes))
qqline(resid(fixed_slopes), col = 2,lwd=2,lty=1)
qqnorm(u)
qqline(u, col = 2,lwd=2,lty=1)

# visualise
compare.fits(AgeFirstKill~Sex | Motive, data=df, fixed_slopes, baseline,
             clusters=2)

# model 2 (Sex is too imbalanced for VCM so let's try Race)
baseline = lmer(AgeFirstKill~1 +(1|Motive), data=df)
fixed_slopes = lmer(AgeFirstKill~Race +(1|Motive), data=df,REML=F)
summary(fixed_slopes)
tab_model(fixed_slopes)

```

```

# assumption checking
ranomeffects <- ranef(fixed_slopes)
u <- ranomeffects[["Motive"]][["(Intercept)"]]

par(mfrow = c(1,2))
par(mar = c(4, 4, 4, 4))
qqnorm(resid(fixed_slopes))
qqline(resid(fixed_slopes), col = 2,lwd=2,lty=1)
qqnorm(u)
qqline(u, col = 2,lwd=2,lty=1)

# visualise
compare.fits(AgeFirstKill~Race | Motive, data=df, fixed_slopes, baseline,
             clusters=2)

# 3.4) Random Slopes model-----
rm(list=setdiff(ls(), "df"))
dev.off()

# model
baseline = lmer(AgeFirstKill~1 +(1|Motive), data=df, REML=F)
fixed_slopes = lmer(AgeFirstKill~Race +(1|Motive), data=df,REML=F)
random_slopes = lmer(AgeFirstKill~Race +(Race|Motive), data=df,REML=F)
summary(random_slopes)
icc(baseline)
icc(fixed_slopes)
icc(random_slopes)

model.comparison(baseline,fixed_slopes)
model.comparison(baseline,random_slopes)
model.comparison(fixed_slopes,random_slopes)

# assumption checking
ranomeffects <- ranef(random_slopes)
u0 <- ranomeffects[["Motive"]][["(Intercept)"]]
u1 <- ranomeffects[["Motive"]][["RaceWhite"]]

par(mfrow = c(2,2))
par(mar = c(4, 4, 4, 4))
qqnorm(resid(random_slopes),main="")
qqline(resid(random_slopes), col = 2,lwd=2,lty=1)
qqnorm(u0,main="")
qqline(u0, col = 2,lwd=2,lty=1)
qqnorm(u1,main="")
qqline(u1, col = 2,lwd=2,lty=1)

# visualise
compare.fits(AgeFirstKill~Race | Motive, data=df, fixed_slopes, random_slopes,
             clusters=5)

#=====
# step 4: Model Comparison-----

```

[illegible]





```

        profile$State=="WI" |
        profile$State=="IL" |
        profile$State=="MI" |
        profile$State=="IN" |
        profile$State=="OH" |
        profile$State=="NE", "Midwest",
        ifelse(profile$State=="AR" |
                profile$State=="LA" |
                profile$State=="MS" |
                profile$State=="AL" |
                profile$State=="TN" |
                profile$State=="GA" |
                profile$State=="KY" |
                profile$State=="FL" |
                profile$State=="NC" |
                profile$State=="SC" |
                profile$State=="VA" |
                profile$State=="DC" |
                profile$State=="DE" |
                profile$State=="WV", "Southeast",
                ifelse(profile$State=="NA", "NA",
                        "Northeast")))))

profile$Region <- as.factor(profile$Region)

# 5.2) Modeling-----
rm(list=setdiff(ls(), c("df", "profile")))
# 5.2.1) Univariate Binary Response Model-----
# model
profile_state <- profile[order(profile$State),]

F0 <- logit(Race) ~ 1 + Whitevictims + (1 | State)
UniModel1 <- runMLwiN(Formula=F0,D="Binomial",data=profile_state,
                      estoptions = list(resi.store=T))
summary(UniModel1)

# residual checking
state_residuals <- na.omit(UniModel1$residual[["lev_2_resi_est_Intercept"]])
state_residuals_sd <- na.omit(UniModel1$residual[["lev_2_resi_var_Intercept"]]**0.5)
pair <- data.frame(state_residuals, state_residuals_sd)
pair <- pair[order(pair$state_residuals, decreasing = TRUE), ]
qqnorm(state_residuals)
qqline(state_residuals, col = 2, lwd=2, lty=1)

par(lwd = 1, cex=1)
x <- as.vector(summary(na.omit(profile_state$State)))
y <- table(na.omit(profile_state$State))
xx <- barplot(sort(x, decreasing = TRUE), axisnames = T,
               names.arg=names(sort(y, decreasing = TRUE)), xlab= "State",
               ylab="Serial Killer Sample Size", space=0, las=2)
text(x = xx, sort(x, decreasing = TRUE), label = sort(x, decreasing = TRUE),
     pos = 3, cex = 1, col = "red", srt=90)

```

```

Indx <- seq(1,dim(pair)[1])
plot(Indx,pair$state_residuals,pch=19,col="blue",
     main="",xlab="State Index", ylab="Residual Value", ylim=c(-1,1.1))
arrows(x0=Indx, y0=pair$state_residuals-pair$state_residuals_sd, x1=Indx,
       y1=pair$state_residuals+pair$state_residuals_sd,
       code=3, angle=90, length=0.05)

# 5.2.2) Multivariate Binary Response Model-----
rm(list=setdiff(ls(), c("df","profile","profile_state")))
# model
F1 <- c(logit(Race),logit(Sex),logit(Motive)) ~ 1 + (1 | State)
MultModel1 <- runMLwiN(Formula=F1, D=c("Mixed","Binomial","Binomial","Binomial")
, data=profile_state, estoptions = list(Meth=0))
summary(MultModel1)

# 5.2.3) Clustering By Region-----
profile_region <- profile[order(profile$Region),]

F2 <- c(logit(Race),logit(Sex),logit(Motive)) ~ 1 + (1 | Region)
MultModel2 <- runMLwiN(Formula=F2, D=c("Mixed","Binomial","Binomial","Binomial")
, data=profile_region, estoptions=list())
summary(MultModel2)

# 5.2.4) Adding predictor variables-----
F3 <- c(logit(Race),logit(Sex),logit(Motive)) ~ 1 + RapedVictims +
KilledWithGun + BoundVictims + (1 | State)
MultModel3 <- runMLwiN(Formula=F3, D=c("Mixed","Binomial","Binomial","Binomial")
, data=profile_state, estoptions = list(Meth=1))
summary(MultModel3)

```