Modeling Sleep Quality Using Mixed Models for Ordinal Dependent Data

Giorgio Bonetta (25-325-200)
Maximilian Brunner (21-333-125)
Simon Derleth (22-415-871)

Linear Model for Dependent Data
Master of Science in Statistics
Geneva School of Economics and Management
University of Geneva

Lecturer: Romain Pic
Teaching Assistant: Dongmei Liu

Fall 2025

Contents

UNIVERSITÉ DE GENÈVE

1 Abstract

In this project, we investigate how lifestyle factors and demographic characteristics shape sleep quality in a clustered health dataset. The response variable, Quality of Sleep, is measured on an ordinal scale, and observations are grouped by occupation. We begin by fitting standard linear and mixed-effects models to highlight their limitations when applied to ordinal outcomes and clustered data. Diagnostic results reveal strong violations of normality in linear models, as well as substantial occupation-level heterogeneity that requires a mixed-effects structure. We then estimate a Cumulative Link Mixed Model (CLMM) to appropriately handle both the ordinal scale and the hierarchical structure of the data. The CLMM identifies stress level as the strongest negative predictor of sleep quality, while physical activity, age, and gender exhibit positive associations. Finally, after a Monte Carlo analysis, we conclude that Mixed effects greatly improve the performance of a model as they account for group-level heterogeneity, and that although CLMM offers the most coherent statistical framework for ordinal data, it can empirically fail to provide stable results when met with finite sample sizes and unbalanced group structures that hinder the convergence of complex maximum likelihood.

2 Dataset Description

The Sleep Health and Lifestyle dataset consists of simulated data reflecting realistic relationships between sleep quality, stress, physical activity, age, occupation, and physiological indicators. Despite being synthetic, the dataset mimics typical patterns observed in health studies. The dataset contains 374 individuals described by 13 variables. Table 1 summarizes the variables used in the analysis, including their type, range, and a short description of their role in the modeling framework. Notice that the dataset contains "Normal" and "Normal Weight" categories. Due to a lack of interpretability and to not arbitrarily remove information, we decided not to merge those factor levels.

Variable	**Type**	**Description**
Gender	Categorical	Male / Female
Age	Continuous	Centered and scaled age of individual
Occupation	Categorical	11 occupational groups (cluster variable)
Sleep Duration	Continuous	Hours of sleep per night (scaled)
Physical Activity Level	Continuous	Self-reported activity score (scaled)
Stress Level	Continuous	Self-reported stress score (scaled)
Daily Steps	Continuous	Number of steps per day (scaled)
Heart Rate	Continuous	Resting heart rate (scaled)
BMI Category	Categorical	Normal / Normal Weight / Overweight / Obese
Sleep Disorder	Categorical	None / Apnea
Quality of Sleep	Ordinal	Response: categories 4–9

Table 1: Summary of variables used in the analysis.

3 Research Question and Methodology

Understanding the determinants that affect sleep quality is a key topic in public health research. Today's lifestyle and work environments create significant differences in how people sleep. This variation makes it inherently complex to analyze the factors that influence sleep. The aim of this study is to evaluate how demographic characteristics influence Quality of Sleep relative to daily behaviours, while accounting for heterogeneity across occupational groups and for the ordinal nature of the response variable. More specifically, we ask:

> *How do daily-related behaviour (Sleep Duration, Physical Activity Level, Daily Steps, Stress level) influence sleep quality compared to demographic structural attributes (Gender, Age) and to what extent the impact of stress varies across occupational groups, once the clustered data structure is taken into account?*

To answer this question, we model a strategy that handles the clustered structure of the dataset and the ordinal nature of the response variable. Finally, we conduct a Monte Carlo analysis to evaluate the convergence characteristics of each model.

4 Linear Regression Model

First, we fit a full linear regression model that includes every variable and check multicollinearity using the Generalized Variance Inflation Factor (GVIF), introduced in Fox and Monette (1992). According to Kutner et al. (2005), if the square of the scale GVIF for a variable is higher than 10, then it should be removed from the model. Failure to remove the variable may artificially inflate standard errors, masking the true significance of the predictor. The variable Blood Pressure is dropped as it is perfectly collinear. From Table 2 we observe that Stress Level, Age, and Heart rate have a squared scaled GVIF greater than 10. We do not remove these variables as we can observe whether adding random effects can improve this metric. We now fit a standard multiple linear regression model in which the response variable, Quality of Sleep, is treated as a numeric score. The model is specified as :

$$\begin{aligned}
\text{Quality of Sleep}_i = \beta_0 &+ \beta_1 \text{Gender}_i + \beta_2 \text{Age}_i + \beta_3 \text{Occupation}_i + \beta_4 \text{Sleep Duration}_i \\
&+ \beta_5 \text{Physical Activity Level}_i + \beta_6 \text{Stress Level}_i + \beta_7 \text{BMI Category}_i \\
&+ \beta_8 \text{Heart Rate}_i + \beta_9 \text{Daily Steps}_i + \beta_{10} \text{Sleep Disorder}_i + \varepsilon_i,
\end{aligned}$$

with $i \in \{1, \ldots, 374\}$ representing the individuals and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ in the standard linear regression specification, which may not hold true with an ordered response variable. Matrix (E.1) shows the matrix structure.

4.1 Diagnostic Assessment of the Linear Regression Model

Table 3 reports an adjusted $R^2$of 0.965, which is expected given that the response variable was synthetically generated from a linear combination of predictors with added noise. In real ordinal data, such values would be uncommon because categorical grouping typically reduces explained variance. Table 4 shows that standard errors vary across occupations, reflecting the unbalanced group structure. Unequal group sizes not only influence standard errors but also challenge the OLS assumption of homoskedasticity, as estimates from small groups tend to be unstable. This issue is confirmed in Figure A.4, which displays clear occupation-specific residual patterns, indicating clustering. Such dependence violates homoskedasticity and leads to underestimated standard errors and inflated significance levels (Moulton 1990), consistent with the large t-values in Table 3. As noted by Cnaan, Laird, and Slasor (1997), linear mixed models are better suited for unbalanced grouped data due to partial pooling (Gelman and Hill 2007). Treating each occupation as a separate fixed effect in a simple linear regression consumes many degrees of freedom and risks overfitting, whereas a random-effects formulation captures group variability more parsimoniously by estimating only their variance. Further diagnostics indicate additional violations of OLS assumptions. Figure A.1 shows discretization in the error term due to the ordinal nature of the outcome, contradicting the assumption of continuous random noise. The QQ-plot (Figure A.2) confirms deviations from normality, and the residual histogram (Figure A.3) reveals right skewness and excess kurtosis. Although the linear model suggests that Stress Level and Sleep Duration, along with demographic predictors such as Gender and Age, are associated with Sleep Quality, these conclusions are unreliable. Residuals are non-normal and occupation-dependent, and standard errors are biased. Consequently, the linear specification provides only a limited and potentially misleading representation of the relative importance of lifestyle and demographic factors. These diagnostic limitations motivate the transition to mixed-effects models.

5 Linear Mixed Model

As noted in the limitations of linear regression, the Sleep Health and Lifestyle dataset exhibits a clustered structure based on occupation. Individuals within the same profession are likely to share work-related conditions such as schedules, environmental stressors, shift patterns, and daily behavioural constraints, creating intra-group correlation. This dependence violates the independence assumption of OLS estimation. Ignoring clustering typically leads to underestimated standard errors and inflated Type I error rates, which can result in spurious findings (Moulton 1990). Linear mixed-effects models can partially address this issue by introducing random intercepts or slopes to capture occupation-level heterogeneity. This approach allows the baseline sleep quality to differ across professional groups and appropriately models within-group dependence.

Type I error: incorrectly rejecting a true null hypothesis. The risk of making this error is denoted by alpha a
5% in studies

5.1 Linear Mixed Model with Random Intercept

We estimate a linear mixed-effects model with a random intercept for occupation in order to account for the group-based organization of the data. Matrix (E.2) represents the structure of the fixed and random effects. :

$$\begin{aligned}\text{Quality of Sleep}_{ij} = \beta_0 &+ \beta_1 \text{Gender}_{ij} + \beta_2 \text{Age}_{ij} + \beta_3 \text{Sleep Duration}_{ij} + \beta_4 \text{Physical Activity Level}_{ij} \\ &+ \beta_5 \text{Stress Level}_{ij} + \beta_6 \text{BMI Category}_{ij} + \beta_7 \text{Heart Rate}_{ij} + \beta_8 \text{Daily Steps}_{ij} \\ &+ \beta_9 \text{Sleep Disorder}_{ij} + u_{0j} + \varepsilon_{ij},\end{aligned}$$

where $i \in \{1,\ldots,374\}$ indexes individuals, $j \in \{1,\ldots,11\}$ indexes occupations, $u_{0j} \sim \mathcal{N}(0,\sigma_u^2)$ represents the random intercept capturing occupation-specific baseline differences, and $\varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2)$ denotes the individual-level residual error. The introduction of a random intercept explicitly separates between-group and within-group sources of variation. This prevents occupational heterogeneity from being mistakenly absorbed by the fixed effects, thereby producing a more realistic covariance structure. The estimated variance of the random intercept ($\hat{\sigma}_u^2 = 0.122$) indicates substantial differences in baseline sleep quality across professions—heterogeneity that the OLS specification cannot accommodate. The diagnostic plots further illustrate the improvement achieved by accounting for occupation-level variation. The Residuals vs. Fitted plot (Figure B.4) no longer exhibits the systematic patterns observed in the OLS model, showing that between-occupation differences have been successfully absorbed by the random intercept. Similarly, residual medians across occupations align more closely (Figure B.3), indicating a more homogeneous distribution of errors within groups. The Q–Q plot (Figure B.2) displays a clearer approximation to normality, and the Fitted vs. Observed comparison (Figure B.1) confirms a noticeable reduction in residual variance. Finally, the caterpillar plot of random intercepts (Figure B.5) highlights the differences between occupational groups. These estimates reflect variation in baseline sleep quality across professions, consistent with the random intercept variance reported in Table 6. This confirms that the random-intercept model captures a level of group structure that is invisible to ordinary linear models and essential for accurately modelling sleep quality determinants.

5.2 Linear Mixed Model with Random Slope

We further extend the mixed-effects specification by allowing the effect of Stress Level to vary across occupational groups. Matrix (E.2) represents the structure of the fixed and random effects. This leads to a random-slopes model of the form:

$$\begin{aligned}\text{Quality of Sleep}_{ij} = \beta_0 &+ \beta_1 \text{Gender}_{ij} + \beta_2 \text{Age}_{ij} + \beta_3 \text{Sleep Duration}_{ij} + \beta_4 \text{Physical Activity Level}_{ij} \\ &+ \beta_5 \text{Stress Level}_{ij} + \beta_6 \text{BMI Category}_{ij} + \beta_7 \text{Heart Rate}_{ij} + \beta_8 \text{Daily Steps}_{ij} \\ &+ \beta_9 \text{Sleep Disorder}_{ij} + u_{0j} + u_{1j} \text{Stress Level}_{ij} + \varepsilon_{ij},\end{aligned}$$

where u_{0j} and u_{1j} denote the random intercept and the random slope for stress within occupation j. We assume:

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho\,\sigma_0\sigma_1 \\ \rho\,\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \right), \quad \varepsilon_{ij} \sim \mathcal{N}(0,\sigma^2)$$

which may not hold true because of the ordinal response. This random-slope specification represents a meaningful improvement over both the linear model and the random-intercept LMM. Allowing the effect of Stress Level to vary across occupations captures group heterogeneity that cannot be represented by random intercepts alone. Certain occupational groups (e.g., medical staff) may display heightened sensitivity to stress-induced declines in sleep quality, while others may show relative resilience. The random-slope term incorporates these differences by allowing group-specific deviations from the population-average stress effect. This additional flexibility leads to both statistical and substantive improvements. The residual variance decreases sharply ($\hat{\sigma}^2 = 0.037$ compared with 0.051 in the random-intercept model), and the random-intercept variance is reduced (0.052 versus 0.122), indicating that part of the between-occupation variability is now explained by heterogeneous stress responses. In other words, allowing slopes to vary across groups reallocates some variability from the intercept to the slope dimension, yielding a more accurate representation of occupational structure. The diagnostic plots reinforce this interpretation. The caterpillar plot of random slopes (Figure B.12) reveals substantial heterogeneity in stress sensitivity: some occupations exhibit pronounced declines in sleep quality as stress levels rise, while others show milder responses. The random-intercept caterpillar plot (Figure B.11) and the intercept–slope correlation (Figure B.6) further indicates that occupations with higher baseline sleep quality tend to exhibit weaker stress reactivity. These patterns align with the hierarchical framework of Laird and Ware (1982) and highlight interactions between occupational context and behavioural determinants that cannot be captured by models imposing a single, population-wide stress effect.

5.3 Diagnostic Assessment of the Linear Mixed Models

The VIF analyses in Tables 5–7 confirm a substantial reduction in collinearity once mixed-effects structures are introduced. Notably, including a random slope brings the Stress Level GVIF well below the critical threshold of 10, indicating that the strong dependence between Stress Level and Occupation observed in the OLS model has been properly absorbed by the hierarchical structure. Further evidence from the diagnostic tables (Tables 6–8) and graphical assessments (Figures B.1–B.12) shows that the mixed-effects specifications clearly outperform the OLS model. Figures B.8 show that the residuals are closer to a normal distribution than before. Comparing figures A.4 with B.7, B.3, we observe that the median of residuals gets closer to zero when adding mixed effects and that the heights of the boxplot diminish, which gets us closer to the independence assumption that was violated in the fixed effects regression. Looking at Tables 5–7, we observe a decrease in the estimated fixed effect of Stress Level, from -0.68 in the mixed effects model with random intercept to -0.92 in the mixed effects model with random slope. This means the model with random intercept underestimated the effect of stress because it treated job differences as fixed effects, while the hierarchical model properly separates these different sources of variation. The estimated variance of the random slope for Stress Level (0.230) indicates substantial heterogeneity across occupations. This validates that assuming a constant stress effect across groups is structurally incorrect. Additional differences emerge once group-level variation is properly modeled. For example, Daily Steps becomes statistically significant, whereas the effect of Gender shrinks from 0.56 in the OLS model to 0.23 in the random slope model. This attenuation reflects the strong association between Gender and Occupation ($p < 0.001$), implying that the OLS model partially confounded demographic and occupational structure. The likelihood ratio test in Table 9 favors the random-slope model. Table 10 reports a parametric bootstrap following the recommendations of Halekoh and Højsgaard (2014). Based on 1,000 bootstrap replications, the resulting p-value is < 0.001, providing robust confirmation that variation in the stress effect across occupations is statistically meaningful.

Overall, the diagnostics show that linear mixed-effects models account for occupational clustering, produce more homogeneous residuals, and reveal interpretable patterns of group-level heterogeneity. The mixed-effects models clarify how lifestyle behaviours and demographic factors jointly shape sleep quality once occupational clustering is properly accounted for. Stress Level remains the strongest lifestyle determinant, while Age and Gender retain consistent demographic effects. Importantly, the introduction of random slopes demonstrates that the impact of stress varies substantially across occupations, indicating that structural and environmental constraints meaningfully condition lifestyle effects.

6 Limitations of Linear Models with Ordinal Response and Alternative

Linear models and their mixed extensions rely on assumptions that are incompatible with non-continuous ordered outcome variables. In our dataset, Quality of Sleep is measured on an ordered scale with six categories, producing discrete and non-normal residuals. Treating such outcomes as continuous induces heteroskedasticity and violates linearity and normality assumptions, even when random effects are included. As highlighted by Liddell and Kruschke (2018), applying linear or mixed-effects models to ordinal data often leads to predicted values outside the admissible range, unstable variance estimates, and coefficients that lack a coherent interpretation. Because the ordinal scale does not guarantee equal spacing between categories, a one-unit change cannot be meaningfully interpreted as a constant shift in the underlying construct, undermining the substantive meaning of the estimated slopes. These limitations motivate the use of modelling frameworks specifically designed for ordered categorical data, such as cumulative link models and their mixed-effects extensions.

Given the inadequacy of linear approaches, a coherent statistical analysis must employ statistical methods specifically designed to handle ordered-scale responses. A common practice is to transform the ordinal response variable into a continuous structure, but this strategy requires strong assumptions. Stevens (1946) emphasizes that an assumption of equidistance between consequent categories is needed. Such an assumption cannot be verified in the case of sleep quality rating as a transition from 6 to 7 may not represent the same improvement as a change from 8 to 9. Additionally, Rhemtulla, Brosseau-Liard, and Savalei (2012) argues that the response must contain at least 5-7 categories for a transformation to be successful. Although sleep quality contains 6 categories, it lies at the lower end of the requirements, making the transformation borderline. Finally, Winship and Mare (1984) notes that it must be verified that the ordinal variable is a discrete realization of an unmeasured continuous variable, which may be difficult to be empirically tested. A coherent statistical analysis must therefore rely on models specifically designed for ordinal data while simultaneously accounting for the intra-class structure of the dataset. Cumulative link models, as described in Agresti (2010), model the ordered nature of the response through cumulative probabilities rather than raw category scores. Their mixed-effects extension, the Cumulative Link Mixed Model (CLMM), first formalized in Hedeker and Gibbons (1994), incorporates random effects to accommodate group-level clustering, similarly to the Mixed Linear Models. Instead of treating sleep quality as a continuous score, the CLMM models the cumulative log-odds of reporting a sleep-quality value at or below a given category, which ensures that predicted probabilities respect the ordering of the scale. Additionally, this proportional-odds structure guarantees that the regression coefficient shifts all cumulative logits by the same amount, which allows for an interpretable effect on the entire ordinal scale.

7 Cumulative Link Mixed Model (CLMM)

To properly account for the ordinal structure of the response variable and the hierarchical grouping induced by occupations, we estimate a Cumulative Link Mixed Model (CLMM) with proportional–odds specification and occupation–specific random effects. After a VIF analysis, several variables that are colinear such as Age, Sleep Duration, Heart rate and Daily Steps must be dropped We fit a full model containing the same parameters as in the linear regression and a reduced model, with a selected subset of the p predictors. From Table 11, we observe that the reduced model performs better for both metrics (AIC and BIC), thus leading us to select the reduced model. For individual i in occupation j, the cumulative log-odds of reporting a sleep-quality score less than or equal to category $k \in \{4,\dots,9\}$ are modeled as

$$\log\left(\frac{\Pr(\text{Quality of Sleep}_{ij} \le k)}{\Pr(\text{Quality of Sleep}_{ij} > k)}\right) = \theta_k - \Big(\beta_1\,\text{Gender}_{ij} + \beta_2\,\text{Age}_{ij} + \beta_3\,\text{Activity Level}_{ij} + \beta_4\,\text{Stress Level}_{ij} + \beta_5\,\text{BMI Category}_{ij} + u_{0j} + u_{1j}\cdot\text{Stress Level}_{ij}\Big)$$

where θ_k denote the ordered threshold parameters. The occupation-specific random intercept and random slope for Stress Level follow a bivariate normal distribution,

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{u_0}^2 & \rho\,\sigma_{u_0}\sigma_{u_1} \\ \rho\,\sigma_{u_0}\sigma_{u_1} & \sigma_{u_1}^2 \end{pmatrix}\right),$$

with estimated variances $\hat{\sigma}_{u_0}^2 = 12.97, \hat{\sigma}_{u_1}^2 = 28.36, \hat{\rho} = 0.58$. The large variance of the random intercept indicates strong between-occupation differences in baseline sleep quality, the sizable random slope variance reflects heterogeneous sensitivity to Stress Level across professional groups, and the positive correlation suggests that occupations with higher baseline sleep quality also tend to show a stronger relationship between stress and the latent sleep quality continuum. The fixed-effects estimates (Table 12) reveal that Stress Level is the dominant determinant of sleep quality. Additionally, the reported coefficients, such as Stress Level ($\hat{\beta}_1 = -16.44$), Physical Activity ($\hat{\beta}_2 = 6.35$), Age ($\hat{\beta}_3 = 6.94$), presence of obesity ($\hat{\beta}_5 = -16.71$) and random effect variances are exceptionally large for a logit-scale model, which could be due to artifacts from the synthetic data generation. The estimated thresholds satisfy the required ordering, $\theta_{4|5} < \theta_{5|6} < \theta_{6|7} < \theta_{7|8} < \theta_{8|9}$, ensuring coherence between the latent variable and the observed ordinal outcomes.

Overall, the CLMM theoretically provides the most appropriate specification for this setting: It respects the ordinal nature of the response, accounts for heterogeneity across occupations, and delivers interpretable coefficients that align with the data structure. The substantial random-effect variances confirm that occupational clustering plays a critical role in shaping sleep outcomes, making the mixed-effects ordinal model a necessary refinement over both OLS and linear mixed-effects alternatives.

8 Monte Carlo Simulation

To evaluate the performance of different approaches in a mixed-effect setting, we conduct a Monte Carlo simulation study with 2,000 runs per model. Our main goal is to evaluate the ability of our model estimators to converge to the true parameter and the normality assumption of the error term. We compare three approaches: simple linear model (with OLS and GLS), two linear mixed models, one with random intercept and one with random slope, and finally the two cumulative link mixed models. As aforementioned, our ordinal variables may have a non-linear scale in a linear setting, which would be expressed by a bias in Monte Carlo. However, our Monte Carlo results indicate that the Linear Model is remarkably robust to this misspecification in this specific setting. To investigate further, we compared Ordinary Least Squares (OLS) with Generalized Least Squares (GLS). Since GLS explicitly accounts for heteroskedasticity (Aitken (1935)), a divergence between the two would indicate variance issues. As shown in (Figure D.1), both estimators yield nearly identical predictions. Therefore, we can retain asymptotic unbiasedness, and also find no evidence of severe heteroskedasticity. A clearer view of OLS and GLS similarities can be assessed with the sampling distribution of the coefficient estimate of the model variables (Figure D.7). An additional analysis was conducted with the untransformed mixed models to test their practical applications. Their error term distribution confirmed once again a similar shape as for the simple linear models (Figure D.4).Additionally, both models achieve near-identical asymptotic unbiasedness results through boxplot representation (Figure D.5).

The CLMM simulations showed optimal coverage (93-97%), though this was accompanied by wider confidence intervals and noticeable bias in the estimates and theta thresholds (Figure 13). (Figure D.2) shows the bias spread, which is slightly skewed and not null, and might contain heteroskedasticity. We observe that the CLMM lacks the unbiasedness seen in the Linear Model simulations. This may have induced a bias in estimating the CLMM thresholds. Other assessments could possibly be conducted through more robust confidence intervals, such as with the bootstrap or Wald's test. The Second CLMM model has a very similar bias spread to the first CLMM model (Figure D.3), but a much higher presence of outliers, which could outline the presence of multicollinearity. Further non-normality assessments can be observed on this bias distribution graph D.6a.

8.1 Robustness of Linear Models and Failure of Cumulative Link Mixed Model

Thanks to the Monte Carlo analysis, we observe that linear models seem to be robust to the ordinal nature of the response variable. This could be explained by multiple factors. First, although borderline, Quality of Sleep has more levels than the necessary threshold of 5 categories described in Rhemtulla, Brosseau-Liard, and Savalei (2012) when treating it as continuous. This could mean that it may not be an issue to treat the ordinal response as continuous . Additionally, the robustness of linear models could mean that the true, unseen distance between subsequent categories may be equal. This could be explained by the synthetic nature of the data, as the "human/psychological" bias when rating sleep quality may not be modeled by the computer when creating the dataset. This means that the computer may instinctively produce a response variable that is equidistant, which is a phenomenon not necessarily seen when measuring human data. While linear models are robust, the CLMM recovers unbiased estimates. This may be caused by various reasons, such as the unbalance of observations when estimating the boundaries. Indeed, CLMM generally requires a sufficient number of observations to stably estimate threshold parameters. Additionally, the low amount of observations in the dataset may cause the Laplace approximation of the Maximum Likelihood method to induce bias(Christensen (2011)). Indeed, contrary to Mixed Models, which were fitted using the REML and Linear Models with OLS (Which yield the same results as ML under the standard assumption of independently distributed errors), CLMM was fitted using the Laplace Approximation of a standard ML (F). Therefore, since the response is of an ordinal nature, and consequently the independently distributed errors assumption is not verified, the OLS methods may yield entirely different results in terms of bias than CLMM.

9 Conclusion

The primary objective of this research was to assess the distinct contributions of lifestyle and demographic factors, of the synthetic Sleep Health and Lifestyle dataset to sleep quality, by addressing the ordinal nature of the outcome variable and the occupation-based clustering. Substantively, we find that that Stress Levels is the predominant factor that explains sleep quality but that it is only significant for the CLMM, and shows that the clustering must be dealt with using a mixed structured model, specifically a Random Slope model, as it gives the best results. The necessity of the Random Slope specification confirms that the impact of stress is not uniform and rather it varies across professions. We also find that BMI levels and fitness-related variables, are strong predictors of sleep quality and that together with Gender have a huge variance in significance between different models. Additionally, we highlighted the trade-off between theoretical coherence and computational stability, as we assessed that although CLMM are theoretically the best models to deal with ordinal response, they may fail when dealing with low-observations, synthetic dataset that model equidistance, likely due to the perfect linearity of the computer-generated data, as they use the Laplace approximation of the Maximum Likelihood estimators which can induce bias. Another alternative is to use Adaptative Gauss-Hermite Quadrature which is recommended in this contexts (Liu and Pierce (1994)). Finally, we conclude that accounting for clustered occupations via random effects greatly improves the estimators' stability, but that the choice for linear and ordinal specifications is nuanced. For synthetic or strictly equidistant data, linear mixed models provide a robust, efficient, and less computationally expensive method. Building on these findings, future investigations should transition to empirical datasets with more observations and where the assumption of equidistance is challenged.

Appendix

A Linear Regression Model

VIF Analysis

Variable	GVIF	Df	$GVIF^{\frac{1}{2 \cdot Df}}$ (Scaled)	Comparable VIF $(Scaled)^2$
Gender	9.8896	1	3.1448	9.89
Age	13.0911	1	3.6182	13.09
Occupation	867.1049	10	1.4025	**1.97**
Sleep Duration	9.7292	1	3.1192	9.73
Physical Activity Level	6.9645	1	2.6390	6.96
Stress Level	24.1431	1	4.9136	**24.14**
BMI Category	62.5786	3	1.9925	3.97
Heart Rate	11.1217	1	3.3349	11.12
Daily Steps	6.2928	1	2.5085	6.29
Sleep Disorder	9.9888	2	1.7778	3.16

Note: 'Comparable VIF' is the squared scaled GVIF, comparable to standard VIF thresholds. Values >10 indicate severe multicollinearity.

Table 2: Generalized Variance Inflation Factors (GVIF) and their scaled equivalents comparable to standard VIF values

Coefficient Estimates

Predictor	Estimate	Std. Error	t-value	$\Pr(>\|t\|)$
Intercept	7.5819	0.0634	119.564	$< 2 \times 10^{-16}$
Gender: Male	0.5646	0.0667	8.466	6.99×10^{-16}
Age	0.4962	0.0328	15.131	$< 2 \times 10^{-16}$
Occupation: Doctor	-0.5959	0.0813	-7.326	1.64×10^{-12}
Occupation: Engineer	-0.7390	0.0748	-9.873	$< 2 \times 10^{-16}$
Occupation: Lawyer	-0.4970	0.0810	-6.133	2.32×10^{-9}
Occupation: Manager	-0.4108	0.2380	-1.726	0.0853
Occupation: Nurse	-0.4073	0.0722	-5.643	3.46×10^{-8}
Occupation: Sales Rep.	-1.6291	0.2052	-7.938	2.79×10^{-14}
Occupation: Salesperson	-0.8773	0.0917	-9.566	$< 2 \times 10^{-16}$
Occupation: Scientist	-0.7026	0.1430	-4.912	1.38×10^{-6}
Occupation: Software Eng.	-0.5353	0.1361	-3.932	0.000102
Occupation: Teacher	-0.5582	0.0684	-8.157	6.17×10^{-15}
Sleep Duration	0.2052	0.0346	5.921	7.62×10^{-9}
Physical Activity Level	-0.0336	0.0307	-1.093	0.2751
Stress Level	-0.6679	0.0448	-14.913	$< 2 \times 10^{-16}$
BMI: Normal Weight	-0.0435	0.0602	-0.722	0.4705
BMI: Obese	-0.2383	0.1684	-1.415	0.1580
BMI: Overweight	-0.5285	0.0737	-7.175	4.34×10^{-12}
Heart Rate	-0.0911	0.0386	-2.364	0.0186
Daily Steps	0.0553	0.0291	1.902	0.0580
Sleep Disorder: None	0.2561	0.0499	5.129	4.82×10^{-7}
Sleep Disorder: Sleep Apnea	0.2446	0.0595	4.110	4.93×10^{-5}
Residual standard error: 0.225 on 351 df				
Multiple $R^2 = 0.9668$, Adjusted $R^2 = 0.9647$				
F-statistic: 463.9 on 22 and 351 df, $p < 2.2 \times 10^{-16}$				

Table 3: Coefficient estimates for the Linear Regression Model treating sleep quality as a numeric outcome.

Standards Errors Comparison by Occupations

Occupation	Standard Error	Relative Uncertainty
Teacher	0.0684	1.00×
Nurse	0.0722	1.06×
Engineer	0.0748	1.09×
Lawyer	0.0810	1.18×
Doctor	0.0813	1.19×
Salesperson	0.0917	1.34×
Software Eng.	0.1361	1.99×
Scientist	0.1430	2.09×
Sales Rep.	0.2052	3.00×
Manager	0.2380	3.48×

Note: 'Relative Uncertainty' is calculated relative to the most precise group (Teachers). The wide gap (3.5x) suggests significant imbalances in group sample sizes.

Table 4: Comparison of Standard Errors by Occupation.

Residual Diagnostics

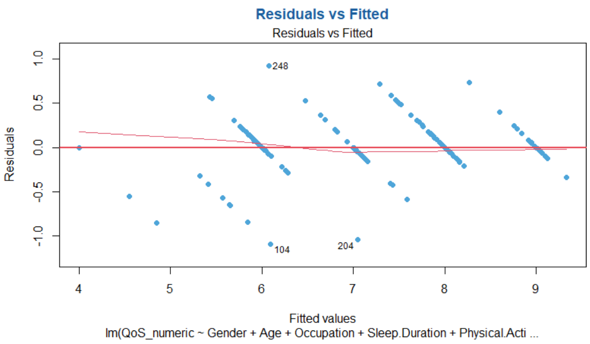


Figure A.1: Residuals vs Fitted for Linear Model

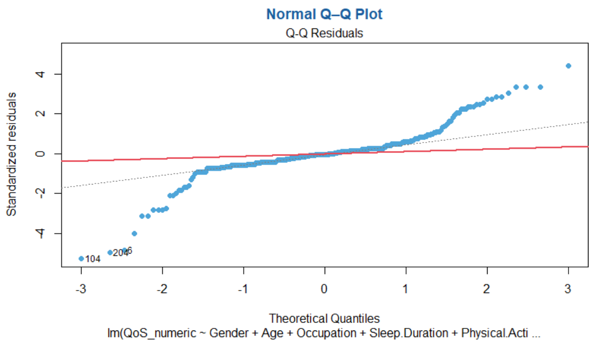


Figure A.2: Normal Q-Q Plot of Residuals

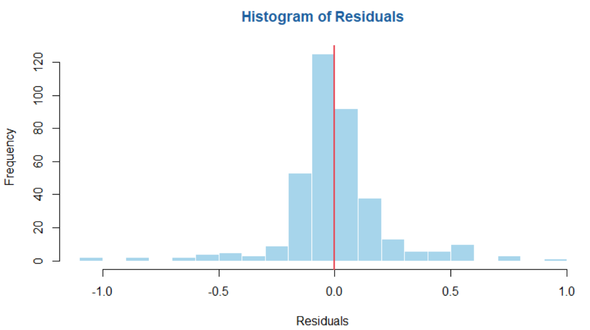


Figure A.3: Histogram of Residuals

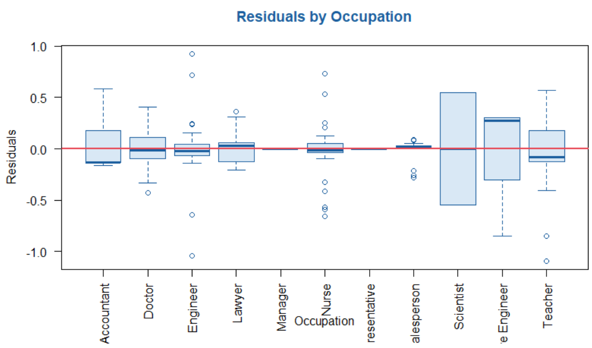


Figure A.4: Residuals by Occupation

B Linear Mixed Models

B.1 Random Intercept Model

VIF Analysis

Variable	GVIF	Df	$GVIF^{1/(2Df)}$ (Scaled)	Comparable VIF $(Scaled)^2$
Gender	2.1164	1	1.4548	2.12
Age	3.2711	1	1.8086	3.27
Sleep Duration	5.5960	1	2.3656	**5.60**
Physical Activity Level	4.5358	1	2.1297	4.54
Stress Level	9.0684	1	3.0114	**9.07**
BMI Category	8.1960	3	1.4199	2.02
Heart Rate	6.9161	1	2.6299	**6.92**
Daily Steps	4.4382	1	2.1067	4.44
Sleep Disorder	1.7303	2	1.1469	1.32

Note: Comparable VIF is the squared scaled GVIF. Values > 10 indicate severe collinearity.

Table 5: GVIF for the Linear Mixed Model with a Random Intercept for Occupation.

Coefficient Estimates

Predictor	Estimate	Std. Error	t-value
(Intercept)	6.982	0.127	54.92
Gender (Male)	0.547	0.064	8.48
Age	0.500	0.032	15.52
Sleep Duration	0.197	0.034	5.72
Physical Activity Level	-0.027	0.031	-0.87
Stress Level	-0.682	0.044	-15.45
BMI: Normal Weight	-0.044	0.060	-0.74
BMI: Obese	-0.329	0.164	-2.01
BMI: Overweight	-0.533	0.072	-7.40
Heart Rate	-0.080	0.038	-2.10
Daily Steps	0.054	0.029	1.86
Sleep Disorder: None	0.252	0.050	5.07
Sleep Disorder: Apnea	0.233	0.059	3.95
Random Intercept Variance	0.122	–	–
Residual Variance	0.051	–	–

Table 6: Coefficient estimates for the Linear Mixed Model with a Random Intercept.

Diagnostic Plots

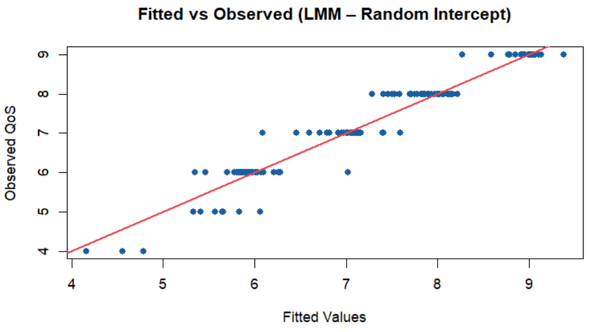


Figure B.1: Fitted vs observed values (Random Intercept).

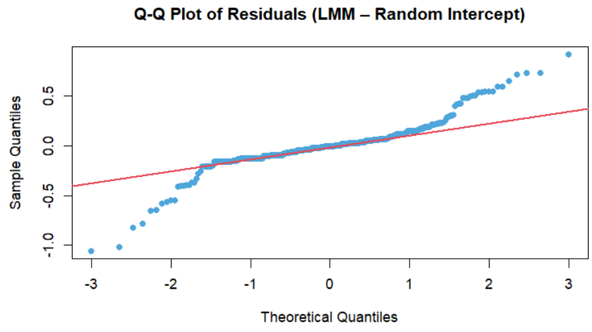


Figure B.2: Q–Q plot of residuals (Random Intercept).

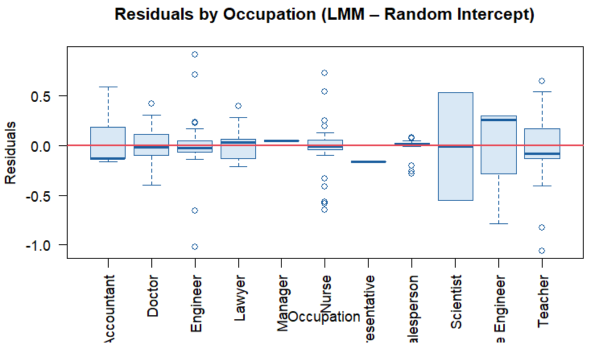


Figure B.3: Residuals by occupation (Random Intercept).

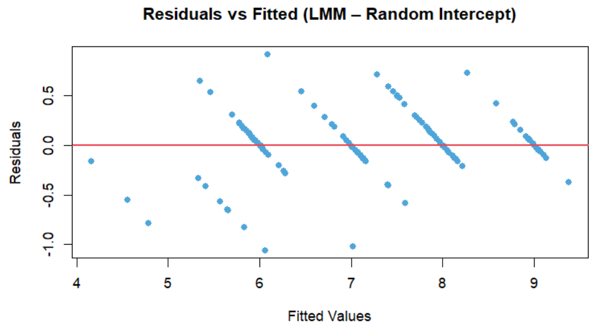


Figure B.4: Residuals vs fitted values (Random Intercept).

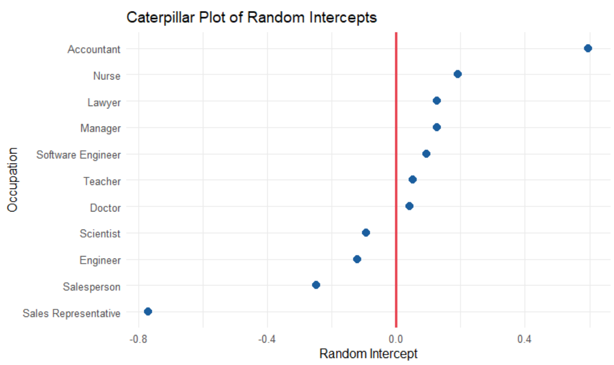


Figure B.5: Caterpillar plot of random intercepts (Random Intercept Model).

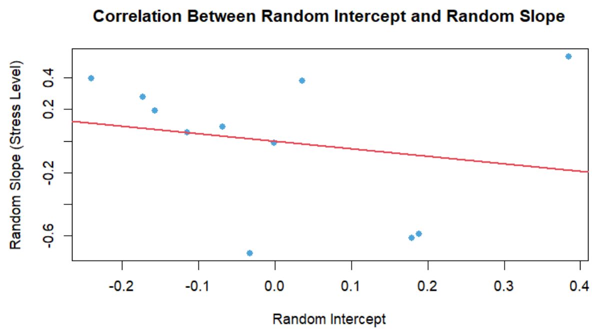


Figure B.6: Intercept–slope correlation between random effects.

B.2 Random Slope Model

VIF Analysis

Variable	GVIF	Df	$GVIF^{1/(2Df)}$	Comparable VIF
Gender	2.0522	1	1.4326	2.05
Age	2.4911	1	1.5783	2.49
Sleep Duration	1.7093	1	1.3074	1.71
Physical Activity Level	3.2479	1	1.8022	3.25
Stress Level	1.0513	1	1.0253	1.05
BMI Category	7.1285	3	1.3873	1.92
Heart Rate	3.7518	1	1.9370	3.75
Daily Steps	3.3473	1	1.8296	3.35
Sleep Disorder	1.5320	2	1.1125	1.24

Table 7: GVIF for the Linear Mixed Model with a Random Slope for Stress.

Coefficient Estimates

Predictor	Estimate	Std. Error	t-value
(Intercept)	7.321	0.103	71.05
Gender (Male)	0.232	0.074	3.13
Age	0.402	0.030	13.46
Sleep Duration	0.133	0.040	3.34
Physical Activity Level	0.044	0.032	1.40
Stress Level	-0.918	0.160	-5.74
BMI: Normal Weight	-0.084	0.060	-1.39
BMI: Obese	0.054	0.153	0.35
BMI: Overweight	-0.563	0.069	-8.12
Heart Rate	-0.065	0.035	-1.87
Daily Steps	0.097	0.026	3.72
Sleep Disorder: None	0.248	0.045	5.54
Sleep Disorder: Apnea	0.252	0.054	4.70
Random Intercept Variance	0.052	–	–
Random Slope Variance (Stress)	0.230	–	–
Corr(Intercept, Slope)	-0.23	–	–
Residual Variance	0.037	–	–

Table 8: Coefficient estimates for the Random-Slope Linear Mixed Model for Sleep Quality.

Diagnostic Plots

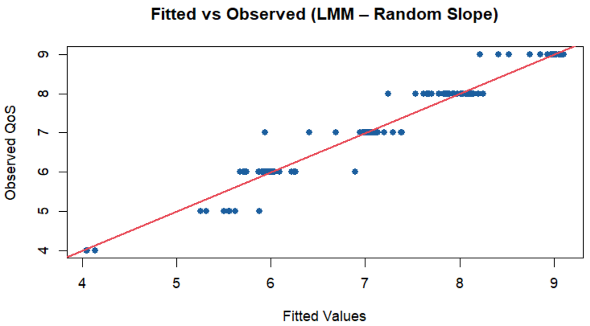


Figure B.7: Fitted vs Observed Values (Random Slope Model).

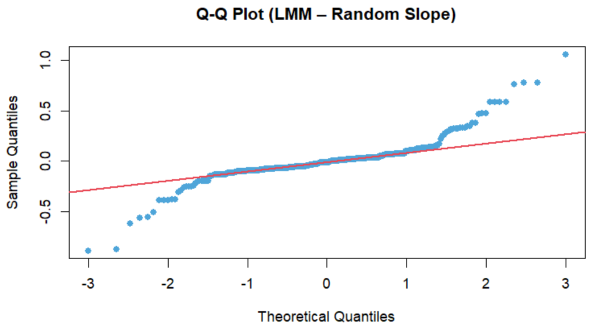


Figure B.8: Q–Q Plot of Residuals (Random Slope Model).

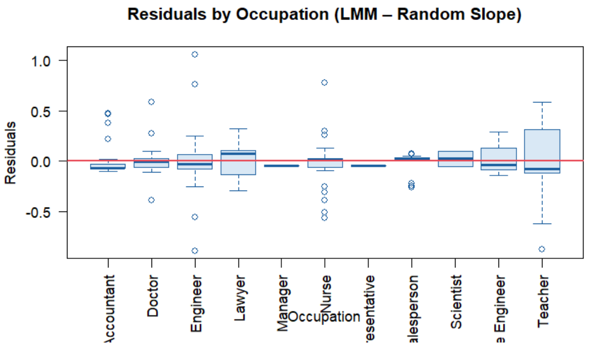


Figure B.9: Residuals by Occupation (Random Slope Model).

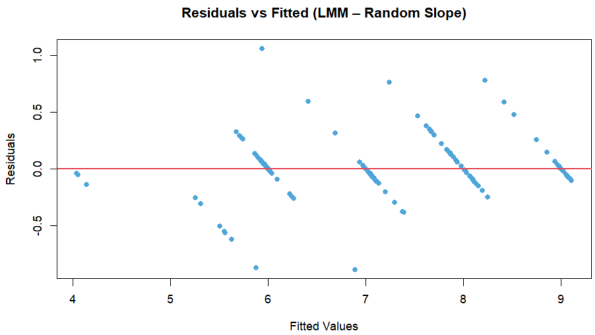


Figure B.10: Residuals vs Fitted (Random Slope Model).

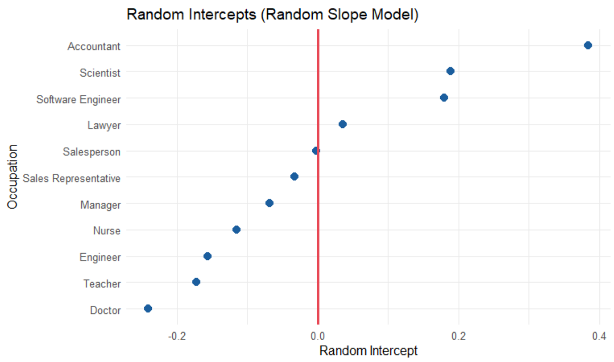


Figure B.11: Caterpillar Plot of Random Intercepts (Random Slope Model).

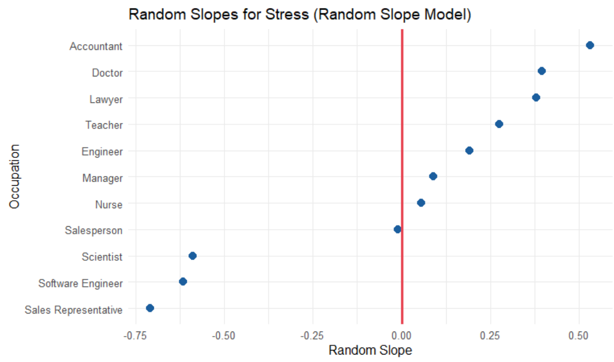


Figure B.12: Caterpillar Plot of Random Slopes for Stress Level.

Likelihood Ratio Test For Mixed Model

Model	**Df**	**LogLik**	**ΔDf**	χ^2	**p-value**
Model 1: Random Intercept	15	-17.306	–	–	–
Model 2: Random Slope (Stress)	17	27.567	2	89.747	$< 2.2 \times 10^{-16}$

Model 2 adds a random slope for Stress Level nested within Occupation.

Table 9: Likelihood ratio test comparing random-intercept and random-slope specifications.

Parametric Bootstrap for Likelihood Ratio Test

Test Method	Statistic	df	p-value
Standard LRT	85.703	2	$< 2.2 \times 10^{-16}$
Parametric Bootstrap Test	85.703	–	0.001

Bootstrap based on 1000 simulations (998 converged).

Table 10: Parametric bootstrap test for the significance of the random slop

C Cumulative Link Mixed Model (CLMM)

Model Selection

Model	Df	AIC	BIC
Reduced Model	15	**134.82**	**193.69**
Full Model	20	134.90	213.38

Note: Lower values indicate better model fit. Best fit is bolded.

Table 11: Model Fit Comparison (AIC & BIC)

Coefficient Estimates

Fixed Effect	Estimate	Std. Error	z-value	p-value
Stress Level	-16.4386	3.6698	-4.479	< 0.001
Physical Activity Level	6.3465	1.7342	3.660	< 0.001
Age	6.9352	1.1910	5.823	< 0.001
BMI: Normal Weight	0.2493	1.9569	0.127	0.899
BMI: Obese	-10.7542	3.0282	-3.551	< 0.001
BMI: Overweight	-12.2419	3.1044	-3.943	< 0.001
Gender: Male	0.8014	2.3110	0.347	0.729
Random Effects (Occupation)	**Variance**	**Std. Dev.**	**Corr**	
Intercept u_{0j}	12.97	3.602	–	–
Slope (Stress Level)	28.36	5.325	0.580	–
Thresholds	**Estimate**	**Std. Error**	**z-value**	
$\theta_{4\|5}$	-45.208	9.373	-4.823	–
$\theta_{5\|6}$	-38.578	7.797	-4.948	–
$\theta_{6\|7}$	-25.918	5.905	-4.389	–
$\theta_{7\|8}$	-2.358	2.588	-0.911	–
$\theta_{8\|9}$	17.339	3.598	4.819	–
Model Fit				
Log-likelihood	-52.41	–	–	–
AIC	134.82	–	–	–

Table 12: Cumulative Link Mixed Model (CLMM) fitted with Laplace approximation

D Monte Carlo Simulations

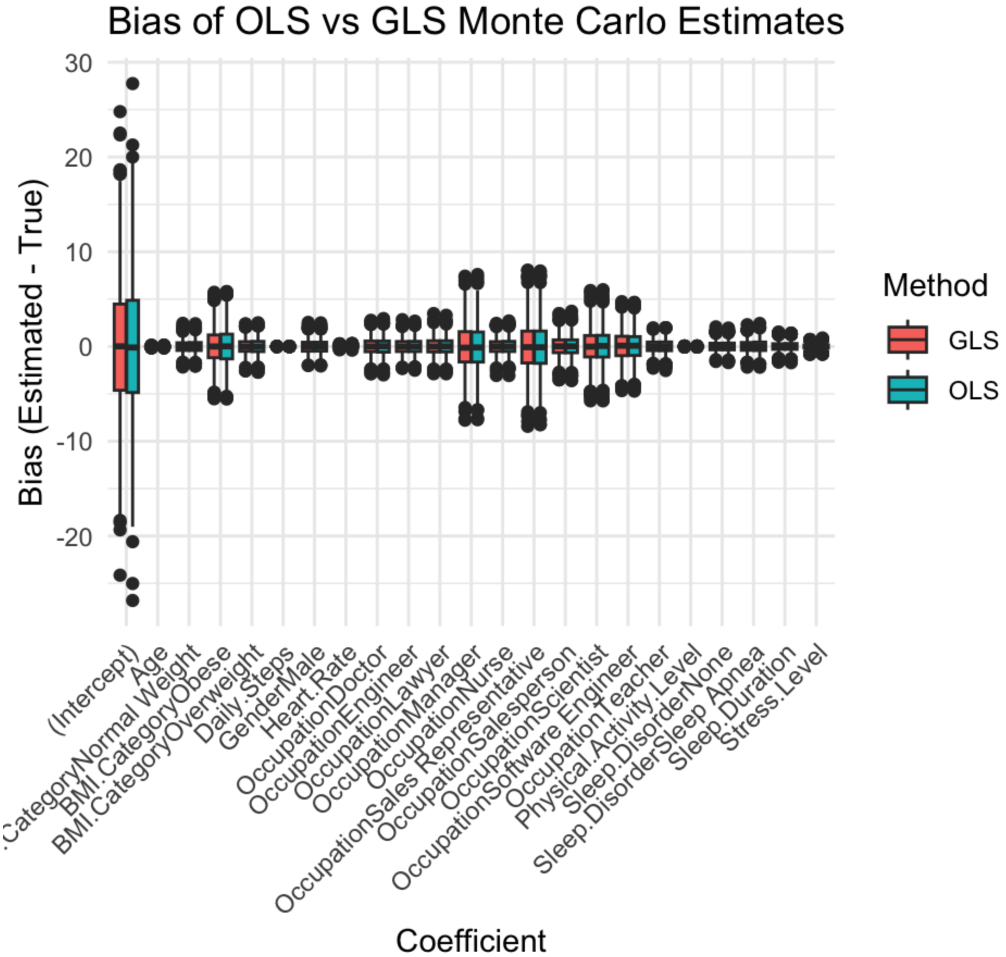


Figure D.1: Distribution of estimation errors for OLS and GLS models.

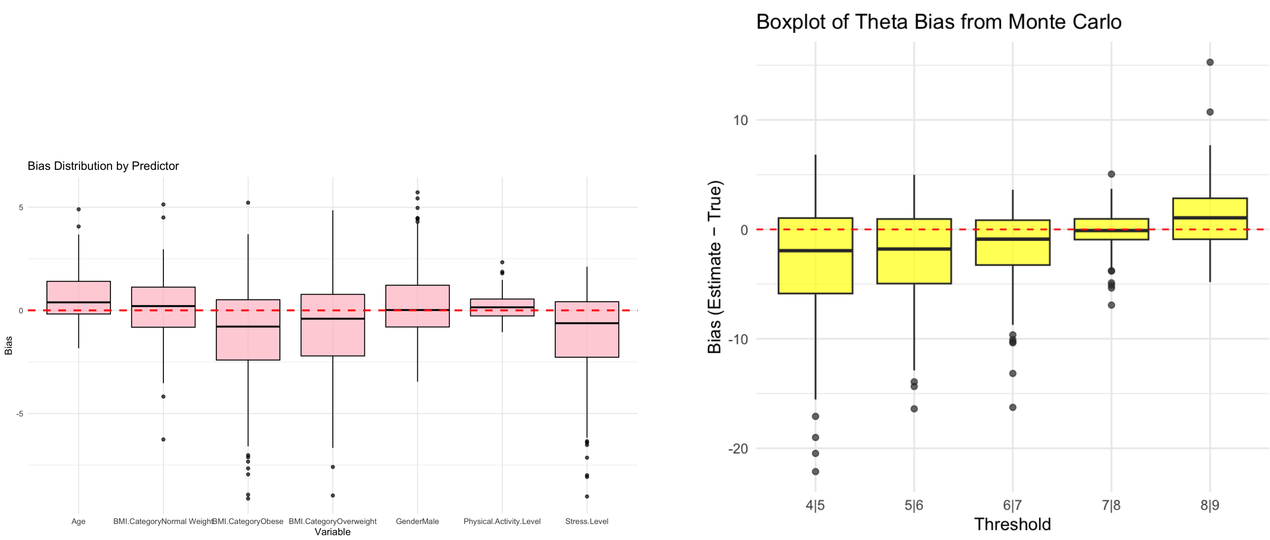


Figure D.2: Distribution of estimation errors for the CLMM with a random intercept.

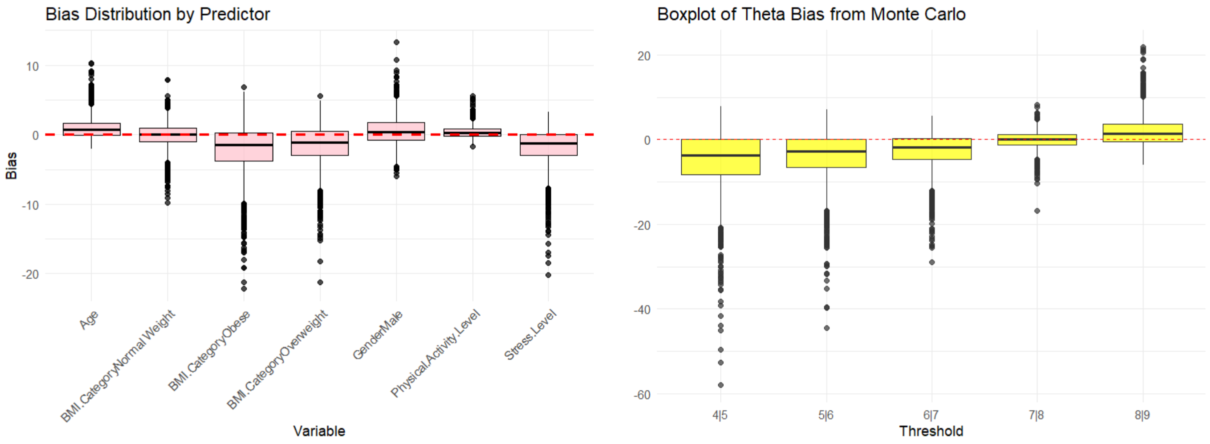


Figure D.3: Boxplots for distribution of estimation errors, CLMM with random slope

Fixed Effects: Bias, RMSE, Coverage

Predictor	True Value	Estimate Mean	Bias	RMSE	95% CI Coverage
Stress	-10.4703	-12.2720	-1.8017	3.2238	0.9523
Activity	1.7238	2.1200	0.3962	0.9326	0.9299
Age	6.4760	7.4203	0.9442	1.7790	0.9523
BMI normal	-0.5094	-0.5964	-0.0870	1.7918	0.9248
BMI obese	-12.2869	-14.3107	-2.0238	4.0711	0.9589
BMI overweight	-9.4970	-11.0268	-1.5298	3.2432	0.9299
Gender male	4.4993	5.0835	0.5842	2.0783	0.8827

Table 13: Summary statistics for CLMM in Monte Carlo simulation

Threshold Estimates: Bias, RMSE, Coverage

Threshold	True Value	Estimate Mean	Bias	RMSE	95% CI Coverage
$\theta_{4\|5}$	-28.2618	-33.2516	-4.9899	8.7899	0.9661
$\theta_{5\|6}$	-23.2084	-27.1615	-3.9532	7.0641	0.9496
$\theta_{6\|7}$	-14.1608	-16.8150	-2.6542	5.1028	0.9419
$\theta_{7\|8}$	-0.3281	-0.4468	-0.1187	2.0825	0.9049
$\theta_{8\|9}$	12.0740	13.9698	1.8959	3.9654	0.9465

Table 14: Summary statistics for threshold estimates in CLMM Monte Carlo simulation

Sampling Distributions and Estimation Errors (mixed effect models intercept vs slope)

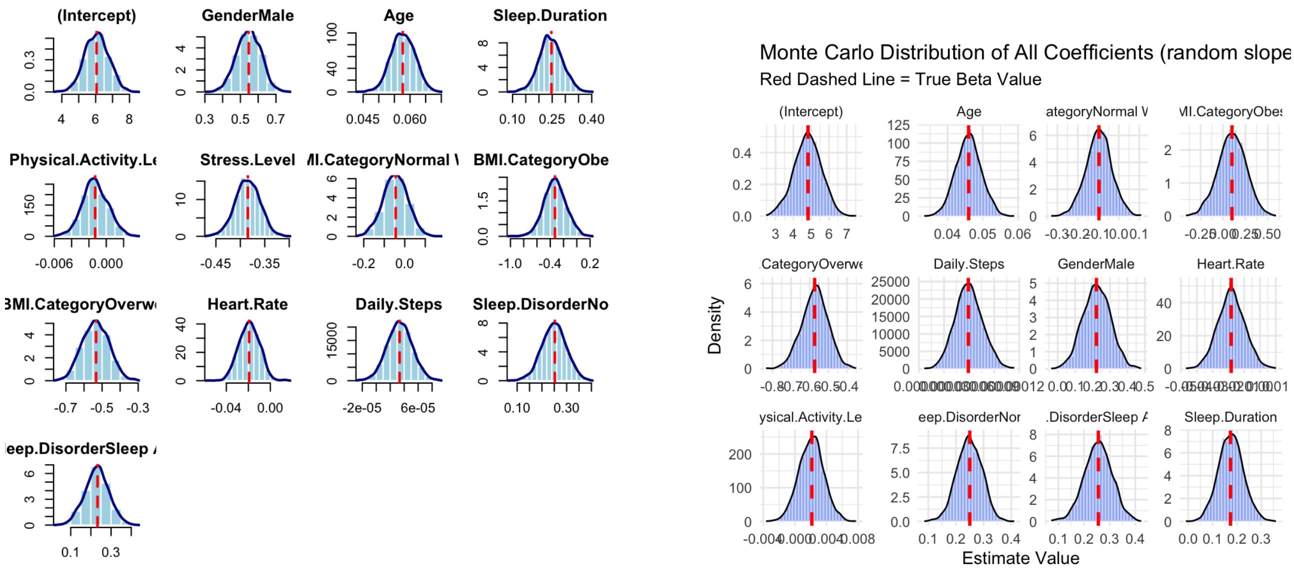


Figure D.4: Sampling distributions of coefficient estimates for random-intercept vs random-slope models.

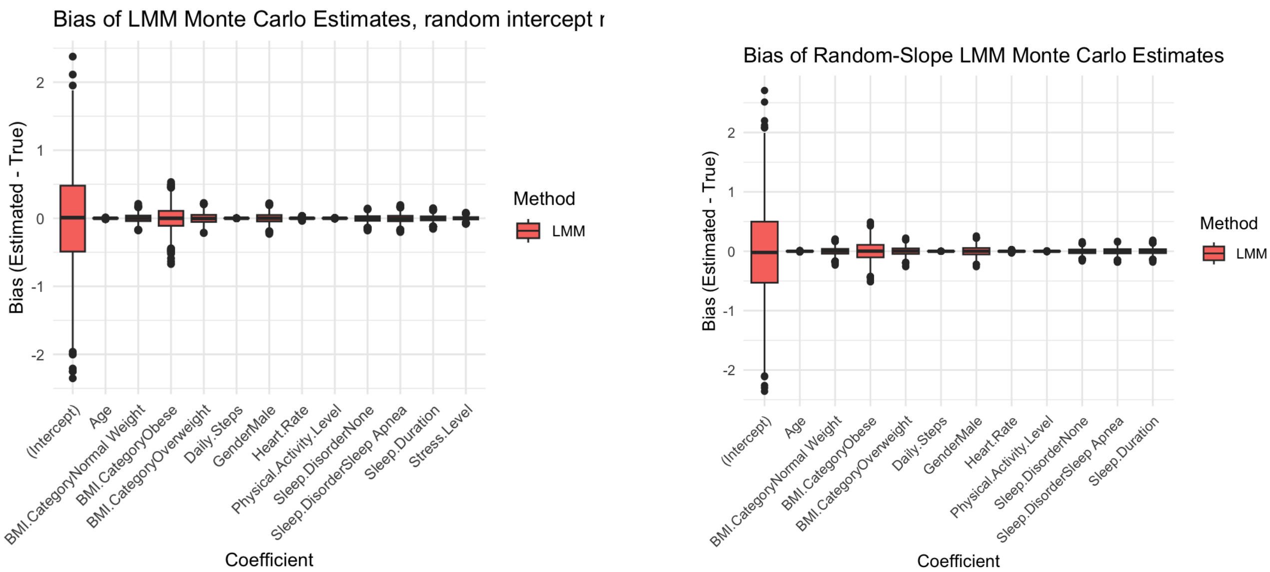
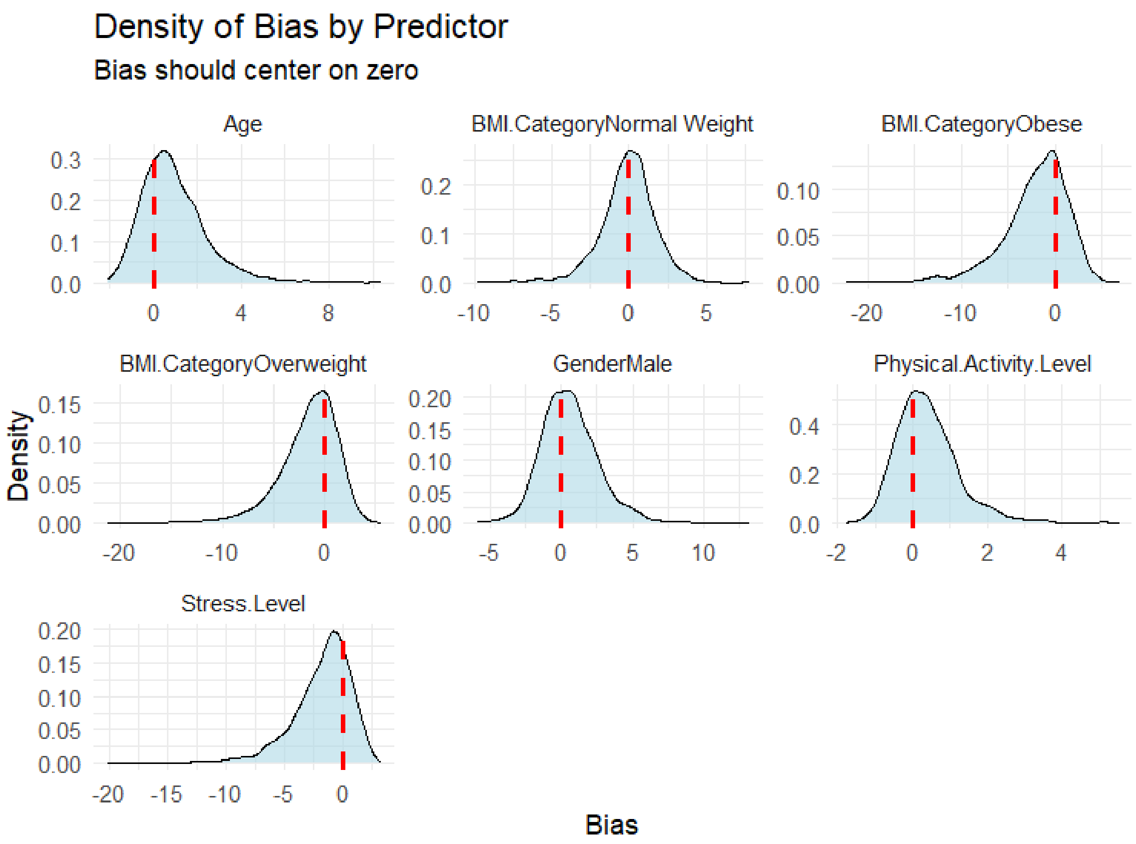


Figure D.5: Distribution of estimation errors (intercept vs slope)

Predictor Bias Distribution



(a) Comparison of the bias distribution of ceofficient estimates for CLMM with random intercept.

Extended Simulation Results

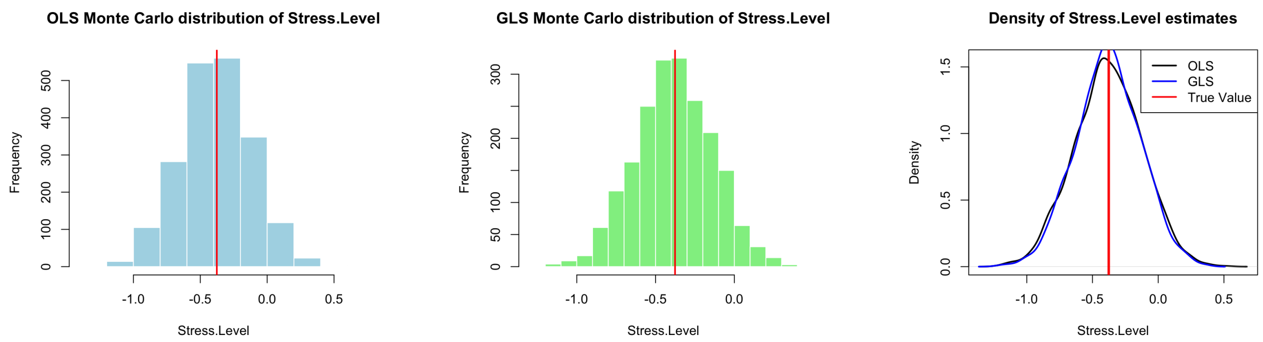


Figure D.7: Comparison of OLS and GLS distribution with a variable of interest

E General Matrix Formulation

E.1 Linear Regression Model

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

Where:

- $\mathbf{y}$ is the 374×1 vector of the response variable (*Quality of Sleep*).
- $\mathbf{X}$ is the 374×23 design matrix for fixed effects.
- β is the 23×1 vector of fixed effects coefficients.
- ε is the 374×1 vector of residual errors.

Fixed Effects Matrix

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_{374} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{374} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 & \text{(Intercept)} \\ \beta_{\text{Gen2}} & \text{(Male)} \\ \beta_{\text{Age}} & \\ \beta_{\text{Occ2}} & \text{(Doctor)} \\ \vdots & (\ldots\text{Other Occs}) \\ \beta_{\text{Occ11}} & \text{(Teacher)} \\ \beta_{\text{Dur}} & \text{(Sleep Dur)} \\ \beta_{\text{Phys}} & \text{(Phys Act)} \\ \beta_{\text{Str}} & \text{(Stress Lvl)} \\ \beta_{\text{BMI2}} & \text{(Normal Wt)} \\ \beta_{\text{BMI3}} & \text{(Obese)} \\ \beta_{\text{BMI4}} & \text{(Overweight)} \\ \beta_{\text{HR}} & \text{(Heart Rate)} \\ \beta_{\text{Step}} & \text{(Daily Steps)} \\ \beta_{\text{Dis2}} & \text{(No Disorder)} \\ \beta_{\text{Dis3}} & \text{(Sleep Apnea)} \end{bmatrix}$$

The Design Matrix $\mathbf{X}$ (abbreviated for width):

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,\text{Gen}} & x_{1,\text{Age}} & x_{1,\text{Occ2}} \cdots x_{1,\text{Occ11}} & x_{1,\text{Dur}} & \cdots & x_{1,\text{BMI2}} \cdots x_{1,\text{BMI4}} & \cdots & x_{1,\text{Dis2}} \cdots x_{1,\text{Dis3}} \\ 1 & x_{2,\text{Gen}} & x_{2,\text{Age}} & x_{2,\text{Occ2}} \cdots x_{2,\text{Occ11}} & x_{2,\text{Dur}} & \cdots & x_{2,\text{BMI2}} \cdots x_{2,\text{BMI4}} & \cdots & x_{2,\text{Dis2}} \cdots x_{2,\text{Dis3}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & x_{n,\text{Gen}} & x_{n,\text{Age}} & x_{n,\text{Occ2}} \cdots x_{n,\text{Occ11}} & x_{n,\text{Dur}} & \cdots & x_{n,\text{BMI2}} \cdots x_{n,\text{BMI4}} & \cdots & x_{n,\text{Dis2}} \cdots x_{n,\text{Dis3}} \end{bmatrix}$$

E.2 Linear Mixed Model - Random Intercept

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$$

Where:

- $\mathbf{y}$ is the 374×1 vector of the response variable (*Quality of Sleep*).
- $\mathbf{X}$ is the 374×13 design matrix for the Fixed Effects.
- β is the 13×1 vector of fixed effects coefficients.
- $\mathbf{Z}$ is the 374×11 design matrix for the Random Effects where (11 Occupations)
- $\mathbf{u}$ is the 11×1 vector of random intercepts.
- ε is the 374×1 vector of residual errors.

Fixed Effects Matrix

$$
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{374} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 & \text{(Intercept)} \\ \beta_{\text{Gen2}} & \text{(Male)} \\ \beta_{\text{Age}} & \\ \beta_{\text{Dur}} & \text{(Sleep Duration)} \\ \beta_{\text{Phys}} & \text{(Physical Activity)} \\ \beta_{\text{Str}} & \text{(Stress Level)} \\ \beta_{\text{BMI2}} & \text{(Normal Weight)} \\ \beta_{\text{BMI3}} & \text{(Obese)} \\ \beta_{\text{BMI4}} & \text{(Overweight)} \\ \beta_{\text{HR}} & \text{(Heart Rate)} \\ \beta_{\text{Step}} & \text{(Daily Steps)} \\ \beta_{\text{Dis2}} & \text{(No Disorder)} \\ \beta_{\text{Dis3}} & \text{(Sleep Apnea)} \end{bmatrix}
$$

The Fixed Design Matrix **X** (13 columns):

$$
\mathbf{X} = \begin{bmatrix} 1 & x_{1,\text{Gen}} & x_{1,\text{Age}} & x_{1,\text{Dur}} & x_{1,\text{Phys}} & x_{1,\text{Str}} & \cdots & x_{1,\text{Dis3}} \\ 1 & x_{2,\text{Gen}} & x_{2,\text{Age}} & x_{2,\text{Dur}} & x_{2,\text{Phys}} & x_{2,\text{Str}} & \cdots & x_{2,\text{Dis3}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,\text{Gen}} & x_{n,\text{Age}} & x_{n,\text{Dur}} & x_{n,\text{Phys}} & x_{n,\text{Str}} & \cdots & x_{n,\text{Dis3}} \end{bmatrix}
$$

Random Effects Matrix

$$
\mathbf{u} = \begin{bmatrix} u_{0,\text{Accountant}} \\ u_{0,\text{Doctor}} \\ \vdots \\ u_{0,\text{Teacher}} \end{bmatrix}, \quad \mathbf{Z} = \begin{array}{c} \left[\begin{array}{cccc} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \dots & 1 \end{array}\right] \end{array} \begin{array}{l} \leftarrow \text{Accountants} \\ \leftarrow \text{Accountants} \\ \\ \leftarrow \text{Doctors} \\ \\ \leftarrow \text{Teachers} \end{array}
$$

E.3 Linear Mixed Model - Random Slope

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \varepsilon$$

Where:

- $\mathbf{y}$ is the 374×1 vector of the response variable (*Quality of Sleep*).
- $\mathbf{X}$ is the 374×12 design matrix for the Fixed Effects.
- β is the 12×1 vector of fixed effects coefficients.
- $\mathbf{Z}$ is the 374×22 design matrix for the Random Effects.
- $\mathbf{u}$ is the 22×1 vector of random effects coefficients (11 occupations $\times$ 2 parameters).
- ε is the 374×1 vector of residual errors.

Fixed Effects Matrix

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{374} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 & \text{(Intercept)} \\ \beta_{\text{Gen2}} & \text{(Male)} \\ \beta_{\text{Age}} & \\ \beta_{\text{Dur}} & \text{(Sleep Duration)} \\ \beta_{\text{Phys}} & \text{(Physical Activity)} \\ \beta_{\text{BMI2}} & \text{(Normal Weight)} \\ \beta_{\text{BMI3}} & \text{(Obese)} \\ \beta_{\text{BMI4}} & \text{(Overweight)} \\ \beta_{\text{HR}} & \text{(Heart Rate)} \\ \beta_{\text{Step}} & \text{(Daily Steps)} \\ \beta_{\text{Dis2}} & \text{(No Disorder)} \\ \beta_{\text{Dis3}} & \text{(Sleep Apnea)} \end{bmatrix}$$

The Fixed Design Matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,\text{Gen}} & x_{1,\text{Age}} & x_{1,\text{Dur}} & \cdots & x_{1,\text{HR}} & x_{1,\text{Step}} & x_{1,\text{Dis2}} \cdots x_{1,\text{Dis3}} \\ 1 & x_{2,\text{Gen}} & x_{2,\text{Age}} & x_{2,\text{Dur}} & \cdots & x_{2,\text{HR}} & x_{2,\text{Step}} & x_{2,\text{Dis2}} \cdots x_{2,\text{Dis3}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & x_{n,\text{Gen}} & x_{n,\text{Age}} & x_{n,\text{Dur}} & \cdots & x_{n,\text{HR}} & x_{n,\text{Step}} & x_{n,\text{Dis2}} \cdots x_{n,\text{Dis3}} \end{bmatrix}$$

Random Effects Matrix

Since we add a random slope to the fixed effects :

$$\mathbf{u} = \left[\begin{array}{c} u_{0,\text{Acct}} \\ u_{1,\text{Acct}} \\ \hline u_{0,\text{Dr}} \\ u_{1,\text{Dr}} \\ \hline \vdots \\ \hline u_{0,\text{Teach}} \\ u_{1,\text{Teach}} \end{array}\right] \begin{array}{l} \leftarrow \text{Random Intercept (Group 1)} \\ \leftarrow \text{Random Slope for Stress (Group 1)} \\ \leftarrow \text{Random Intercept (Group 2)} \\ \leftarrow \text{Random Slope for Stress (Group 2)} \\ \\ \leftarrow \text{Random Intercept (Group 11)} \\ \leftarrow \text{Random Slope for Stress (Group 11)} \end{array}$$

The Random Effects Design Matrix $\mathbf{Z}$ is block-diagonal. Let s_i denote the **Stress Level** of individual i.

UNIVERSITÉ DE GENÈVE

$$\mathbf{Z} = \left[\begin{array}{cc|cc|c|cc} 1 & s_1 & 0 & 0 & \dots & 0 & 0 \\ 1 & s_2 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \hline 0 & 0 & 1 & s_k & \dots & 0 & 0 \\ 0 & 0 & 1 & s_{k+1} & \dots & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline 0 & 0 & 0 & 0 & \dots & 1 & s_{374} \end{array}\right]$$

Note: The blocks correspond to the specific Occupation groups. For any given row, only 2 columns are non-zero: the intercept (1) and the Stress Level (s_i) corresponding to that person's specific occupation.

F Mathematical Details – Likelihood Function for the CLMM

In contrast to the Linear Mixed Model, which relies on the assumption of normally distributed residuals to derive a closed-form likelihood, the Cumulative Link Mixed Model (CLMM), as defined in Hedeker and Gibbons (1994), expresses the likelihood as an integral of the product of individual conditional probabilities over the distribution of the random effects.

Let θ denote the vector of threshold parameters and β the fixed effects. The likelihood function is given by:

$$L(\theta, \beta, \Sigma_u \mid \mathbf{y}) = \int_{\mathbb{R}^J} \left[\prod_{i=1}^{n} P(Y_i = y_i \mid \mathbf{u})\right] f(\mathbf{u} \mid \mathbf{0}, \Sigma_u)\, d\mathbf{u}. \tag{1}$$

The conditional probability of observing category k for individual i is determined by the inverse-logit cumulative distribution function F:

$$P(Y_i = k \mid \mathbf{u}) = F(\theta_k - \mathbf{x}_i^\top \beta - \mathbf{z}_i^\top \mathbf{u}) - F(\theta_{k-1} - \mathbf{x}_i^\top \beta - \mathbf{z}_i^\top \mathbf{u}). \tag{2}$$

Since the integral does not admit a closed-form solution, parameter estimation requires numerical approximation. A common approach is to maximize the log-likelihood using the Laplace Approximation, first introduced in Breslow and Clayton (1993):

$$\ell_{CLMM} \approx -\frac{1}{2} \log|\det(\mathbf{H})| + \log f(\mathbf{y} \mid \hat{\mathbf{u}}) + \log f(\hat{\mathbf{u}}), \tag{3}$$

where $\hat{\mathbf{u}}$ is the mode of the conditional distribution of the random effects and $\mathbf{H}$ is the Hessian matrix evaluated at $\hat{\mathbf{u}}$. It is well known, however, that the Laplace Approximation may lead to biased estimates in small samples.

Use of Generative AI

Generative AI tools (primarily OpenAI's ChatGPT) were used exclusively to improve the clarity and grammatical correctness of the written text. All statistical thinking and interpretation of results, and substantive conclusions were carried out independently by the authors. AI tools were also used in several parts of the R code to correct dysfunctional chunks of code, and to implement code using methods beyond the scope of the class. Finally, Generative AI was used to format tables using LaTeX code.

References

Agresti, Alan (2010). *Analysis of Ordinal Categorical Data*. 2nd. John Wiley & Sons.

Aitken, Alexander C. (1935). “On Least Squares and Linear Combination of Observations”. In: *Proceedings of the Royal Society of Edinburgh* 55, pp. 42–48.

Breslow, Norman E and David G Clayton (1993). “Approximate inference in generalized linear mixed models”. In: *Journal of the American Statistical Association* 88.421, pp. 9–25.

Christensen, Rune Haubo Bojesen (2011). “Ordinal regression models for ordinal data”. PhD Thesis. Copenhagen, Denmark: Department of Mathematical Sciences, University of Copenhagen.

Cnaan, Avital, Nan M Laird, and Peter Slasor (1997). “Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data”. In: *Statistics in medicine* 16.20, pp. 2349–2380.

Fox, John and Georges Monette (1992). “Generalized collinearity diagnostics”. In: *Journal of the American Statistical Association* 87.417, pp. 178–183.

Gelman, Andrew and Jennifer Hill (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Halekoh, Ulrich and Søren Højsgaard (2014). “A Kenward–Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models – The R Package pbkrtest”. In: *Journal of Statistical Software* 59.9, pp. 1–30. DOI: 10.18637/jss.v059.i09.

Hedeker, Donald and Robert D Gibbons (1994). “A random-effects ordinal regression model for multilevel analysis”. In: *Biometrics*, pp. 933–944.

Kutner, Michael H et al. (2005). *Applied Linear Statistical Models*. 5th. Boston: McGraw-Hill Irwin.

Laird, Nan M and James H Ware (1982). “Random-effects models for longitudinal data”. In: *Biometrics*, pp. 963–974.

Liddell, Torrin M and John K Kruschke (2018). “Analyzing ordinal data with metric models: What could possibly go wrong?” In: *Journal of Experimental Social Psychology* 79, pp. 328–338.

Liu, Qing and Donald A. Pierce (1994). “A Note on Gauss-Hermite Quadrature”. In: *Biometrika* 81.3, pp. 624–629. DOI: 10.1093/biomet/81.3.624.

Moulton, Brent R (1990). “An illustration of the pitfall of estimating deterministic coefficients on grouped data”. In: *The Review of Economics and Statistics*, pp. 334–338.

Pic, Romain (2025). “Linear Model for Dependent Data: Lecture Notes”. Lecture Slides. University of Geneva, Fall Semester.

Rhemtulla, Mita, Patriciaé Brosseau-Liard, and Victoria Savalei (2012). “When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods”. In: *Psychological Methods* 17.3, pp. 354–373.

Stevens, Stanley Smith (1946). “On the theory of scales of measurement”. In: *Science* 103.2684, pp. 677–680.

Winship, Christopher and Robert D Mare (1984). “Regression models with ordinal variables”. In: *American Sociological Review*, pp. 512–525.