**PSYR 6003 Assignment 4: Linear Mixed Effects Modeling**

Holly Carbyn

Department of Psychiatry, Dalhousie University

PSYR 6003: Fundamentals of Applied Statistics

Dr. Igor Yakovenko

April 26th, 2024

**Assignment 4: Linear Mixed Effects Modeling**

For our data analysis, we used a linear mixed effect model to evaluate a sample of 263 people that were measured with daily questionnaires once a day for up to 20 days. Data analysis was performed using statistical software R version 4.3.2 (Eye Holes). The sample consisted of participant scores on personability traits such as extraversion (i.e., tendency to experience positive affect and be engaged in the social world) and neuroticism (i.e., tendency to experience negative affect) as well as satisfaction with life. We used the ‘Tidyverse’ package (Wickham et al., 2019) to conduct analyses. Cases with missing values were removed automatically upon running the linear mixed effect model.

**Descriptive Statistics & Bivariate Correlations**

Descriptive statistics and bivariate correlations were calculated by creating a correlation table using the ‘apaTables’ package (Stanley, 2021) for our variables of interest; satisfaction with life coded as ‘swl’ (*M* = 4.43, *SD* = 1.61), extraversion coded as tipm.E (*M* = 4.18, *SD* = 1.52), and neuroticism coded as tipm.N (*M* = 3.49, *SD* = 1.54), as seen in *Table 1*. Bivariate correlations were calculated to examine the relationships among these variables. Satisfaction with life was positively correlated with extraversion (*r* = 0.38, *CI* = [0.35,0.41], *p* < 0.01) and negatively correlated with neuroticism (*r* = -0.45, *CI* = [-0.47, -0.43], *p* < 0.01). Additionally, there was a significant negative correlation between extraversion and neuroticism (*r* = -0.33, *CI* = [-.36, -.31], p < 0.01) as seen in *Table 1*. These findings suggest that satisfaction with life is associated with higher levels of extraversion and lower levels of neuroticism, while extraversion is inversely related to neuroticism.

We next used the package ‘flexplot’ (Phillipp, 2021) to create graphs and plots to check the statistical assumptions for our final model. A Histogram of our residuals showed a normal distribution, so the assumption of normality of residuals is met. The assumption of linearity seems to be met when looking at the residual dependence plot as we see a straight horizontal line along zero. The assumption of homogeneity of variance or ‘homoscedasticity’ could be violated due to the slight negative slope of the S-L plot. However, we ignored this possible slight violation.

**Linear Mixed Effect Model**

We fit and ran a linear mixed effect model using the ‘lme4’ package (Bates, 2015) to see if 1) Extraversion is positively associated with satisfaction with life, 2) Neuroticism is negatively associated with satisfaction with life and 3) If the effects will be similar for both level 1 (within participants over time) and level 2 (between participants).

We first visualized the univariate distributions of our analysis variables by creating graphs and plots using the ‘flexplot’ package (Phillipp, 2021). We see that scores for satisfaction with life have a slightly negatively skewed distribution, extraversion has a normal distribution, and neuroticism has a positively skewed distribution. Next, we fit a baseline model with only the dependent variable (swl) and no predictors. We then calculated the Intraclass Correlation (ICC) of the baseline model to see what proportion of all variability in the model is due to the cluster effects, and consequently, whether it is worth it to run a a linear mixed effect model. We found that 74.2% of the variance was due to “clusters”, which, in this case, refers to the repeated measures all coming from the same people (*ICC =* 0.742). Therefore, we need to run a mixed model as opposed to a regular linear regression model. If we were to run a regular linear regression model, our type 1 error would be inflated similar to artificially increasing our sample size 12-fold (*Design effect* = 12.3).

We used a nested model comparison approach to find the best fitting model as this allowed us to statistically evaluate the necessity of modeling random effects. Because we chose to use nested model comparisons, ‘flexplot’ automatically used Maximum Likelihood Estimation (ML) instead of Restricted Maximum Likelihood Estimation (REML), as REML does not accommodate nested model comparisons (even though it is faster and less biased). The nested model comparisons showed that the full model for extraversion, including random and fixed slopes (*AIC =* 10865.86, *BIC =* 10903.99, *Bayes factor* > 100) was preferable to the reduced model for extraversion with only the fixed slope (*AIC =* 10945.67, *BIC =* 10971.08, *Bayes factor* = 0). We then added in fixed slope for neuroticism and saw that it was better to include neuroticism to explain the variance (AIC = 10498.33, BIC = 10542.8, Bayes factor >100) than to not include it and only use extraversion to explain satisfaction with life (AIC = 10865.86, BIC = 10903.99, Bayes factor = 0). Next, we incorporated random slopes for neuroticism and found that random and fixed slopes (*AIC =* 10364.47, *BIC =* 10428.02, *Bayes factor* > 100) were preferable to only the fixed slope (*AIC* = 10498.33, *BIC* = 10542.81, *Bayes factor* = 0).

The model we tested can be represented by the following equation: swl = *β*00 + *β*01(extraversion) + *β*10(neuroticism) + *β*11(extraversion)(neuroticism) + r0 + r1(extraversion) + r2(neuroticism) + e. The package ‘flexplot’ (Phillipp, 2021) was used to create graphs and plots to check the statistical assumptions for our final model. A Histogram of our residuals showed a normal distribution, so the assumption of normality of residuals is met. The assumption of linearity seems to be met when looking at the residual dependence plot as we see a straight horizontal line along zero. The assumption of homogeneity of variance or ‘homoscedasticity’ could be violated due to the slight negative slope of the S-L plot. However, we ignored this possible slight violation.

The results from our linear mixed effect model show that for every one-point increase in extraversion, satisfaction with life goes up by 0.16 points (*B* = 0.16[*CI* = 0.13, 0.19]). For every one-point increase in neuroticism, satisfaction with life goes down by 0.21 points (*B* = -0.21[*CI* = -0.24, -0.18]) as seen in *Table 2*. Therefore, hypotheses one and two are both supported. Extraversion is positively associated with satisfaction with life (H1) and neuroticism is negatively associated with satisfaction with life (H2). Additionally, the model did not predict any variability of satisfaction with life around the overall mean of satisfaction with life for all participants (*r2* = -0.06). However, we predicted 26% of variability of satisfaction with life day to day around the average for each person (*r2 =* 0.2605) (*Table 2*). Therefore, hypothesis three is not supported as effects were not similar for within participants over time (level 1) and for between different participants (level 2). Overall, the take home message is that higher extraversion scores increases satisfaction with life, whereas higher neuroticism scores decreases satisfaction with life, but only across time points within individual participants.

**References**

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Philipp, M. (2021). flexplot: Graphics package to enhance the ggplot2 plotting system. Retrieved from <https://cran.r-project.org/package=flexplot>

Stanley, D. (2021). apaTables: Create American Psychological Association (APA) style tables. Retrieved from <https://cran.r-project.org/package=apaTables>

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S (4th ed.). Springer.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. <https://doi.org/10.21105/joss.01686>

**Appendices**

Table 1

*Means, standard deviations, and correlations with confidence intervals*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | *M* | *SD* | 1 | 2 |
|  |  |  |  |  |
| 1. swl | 4.43 | 1.61 |  |  |
|  |  |  |  |  |
| 2. tipm.N | 3.49 | 1.54 | -.45\*\* |  |
|  |  |  | [-.47, -.43] |  |
|  |  |  |  |  |
| 3. tipm.E | 4.18 | 1.52 | .38\*\* | -.33\*\* |
|  |  |  | [.35, .41] | [-.36, -.31] |
|  |  |  |  |  |

*Note.* *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval for each correlation. The confidence interval is a plausible range of population correlations that could have caused the sample correlation (Cumming, 2014). \* indicates *p* < .05. \*\* indicates *p* < .01.

Table 2

*Bs, 95% CIs, p- values, ICC, and R2 for the final linear mixed effect model*

|  |  |  |  |
| --- | --- | --- | --- |
|  | **swl** | | |
| *Predictors* | *Estimates* | *CI* | *p* |
| (Intercept) | 4.51 | 4.27 – 4.74 | **<0.001** |
| tipm.E | 0.16 | 0.13 – 0.19 | **<0.001** |
| tipm.N | -0.21 | -0.24 – -0.18 | **<0.001** |
| **Random Effects** | | | |
| σ2 | 0.49 | | |
| τ00 id | 2.02 | | |
| τ11 id.tipm.E | 0.02 | | |
| τ11 id.tipm.N | 0.04 | | |
| ρ01 | -0.41 | | |
|  | -0.32 | | |
| ICC | 0.77 | | |
| N id | 262 | | |
| Observations | 4249 | | |
| Marginal R2 / Conditional R2 | 0.094 / 0.788 | | |