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

Julia Fraiha-Pegado

Department of Psychiatry, Dalhousie University

PSYR 6003: Fundamentals of Applied Statistics and Research Design

Dr. Igor Yakovenko

April 23rd, 2024

**Assignment 4 – Linear Mixed Effects Modeling**

The present data analysis was based on the “PSYR6003.A4.sav” fictional data set created and published by Dr. Igor Yakovenko. This file contained 263 participants and measured personality and satisfaction with life. This data was obtained using a 7-point Likert scale (In which, 1 = “strongly disagree” and 7 = “strongly agree”), for up to 20 days. The data set contained self-report information on satisfaction with life, extraversion and neuroticism. Although other information was included in the data set, we excluded it for the purposes of the current analysis. Interested in the relationship between certain personality traits and satisfaction with life, we cloned the data set from Dr. Yakovenko’s repository to conduct an exploratory analysis.

**Data Initialization & Cleaning Procedure**

We initialized the data set by loading it into the integrated environment RStudio (Version 2023.12.1.402) running on the software R (Version 4.2.3) in the background. We used the package “haven” (Wickham et al., 2019) to upload the file into the directory since it was contained within an “.sav” format. After loading, we put the data into an object so we could examine it.

Before starting the exploratory analysis, we visually inspected the raw data to familiarize ourselves with the data set and ensure we had no missing or extraneous values. To manage the data set, we used the package “tidyverse” (Wickham et al., 2019). The raw data file contained 4252 observations and 263 participants, in which we identified 6 rows containing missing values (NAs). We decided to omit these values, to ensure the accuracy of our results, which decreased our sample to 4246 observations and 262 participants.

Since the raw data set contained several variables that were irrelevant to our research question, we decided to remove them from our analysis. We used the package “dplyr” (Wickham et al., 2019) to select only the measures of 1) satisfaction with life, 2) extraversion, and 3) neuroticism.

**Descriptive Statistics**

To determine the average self-reports across the variables of interest we gathered summary statistics for the measures. We obtained mean and standard deviation values using the package “dplyr” (Wickham et al., 2019). Then, we used the package “apaTable” (Stanley, 2021) to report the descriptive statistics and bivariate correlation values.

We attained the mean (*M*) and standard deviation (*SD*) values for satisfaction with life (M= 4.43 SD= 1.61), extraversion (M= 4.18, SD= 1.52), and neuroticism (M= 3.49, SD= 1.54). According to these reports (Table 1), individuals on average reported moderate (4= “neither agree nor disagree/neutral”) contentment with life, while displaying a moderate level of extraversion and slight negative levels of neuroticism (3= “somewhat disagree”).

Additionally, the bivariate correlations reports (Table 1), revealed satisfaction with life to be positively correlated with extraversion (*r* (262) = .38, *p* <.01). These findings could be explained by the level of social interaction one may have; if an individual is extraverted, they may be more sociable and have a larger support system. Inversely, satisfaction with life was found to be negatively correlated with neuroticism (*r* (262) = -.45, *p* <.01). This suggests that those who experience higher neuroticism are likely to be more preoccupied with life, which in turn causes them to be less satisfied with it. Finally, we found a negative bivariate correlation between extraversion and neuroticism (*r* (262) = -.33, *p* <.01). This suggests that extraverted individuals experience lower levels of neuroticism, which is in line with what we would expect according to the available literature on the topic.

**Hypothesis & Analysis Set-Up**

Interested in the relationship between extraversion and neuroticism in predicting life satisfaction, we decided to create a model including the variables using the package “lme4” (Bates et al., 2015). We hypothesized that 1) extraversion would be positively associated with satisfaction with life, 2) neuroticism would be negatively associated with satisfaction with life, and 3) the effects would be similar for level 1 and level 2 (within and between participants, respectively)

**Univariate Distribution**

To merge the univariate plots of the variables of interest we used the package “patchwork” (Pedersen, 2024). We began by visualizing the univariate plots of the variables of interest (satisfaction with life & extraversion), using the package “flexplot” (Fife, 2022). Extraversion (Figure 1) was revealed to be approximately normally distributed indicating a neutral pattern of response for this variable. Neuroticism, exhibited a right skewed arrangement, indicating negative levels of neuroticism. In contrast, the variable satisfaction with life, presented a left skewed pattern, indicating neutral to positive levels of life satisfaction.

**Data Analysis Strategy**

For the purposes of the current analysis, we chose to use a random effect, which was tested using a nested model comparison of our linear mixed effects model. We used maximum likelihood estimation (ML), to accommodate for the nested model comparisons that were built using the package “lme4” (Bates et al., 2015).

We started investigating our hypothesis by fitting a baseline model, which included our dependent variable “satisfaction with life” but no predictor variables, which can be represented by the following equation: baseline = lmer (SatisfactionWithLife~ 1+ (1|id), data.

We did this to determine if we needed to run a linear mixed model. We first obtained the intraclass correlation (ICC) and measures of effect size for our baseline model. The ICC value for the baseline model revealed that 74% of the variance in the data set was due to clustering, which indicates that we should run a mixed model. Similarly, the design effect shows us that by not using linear mixed models we would be increasing our sample size 12-fold, which then would increase our likelihood of committing a Type I error.

**Model and Diagnostics**

Using the package “flexplot” (Fife, 2022), we compared the fixed and random models. We found the random model to be supported based on its fit indices, such as significant p-value, higher Bayes factor and lower AIC and BIC.

We then fitted our model, in which the outcome variable is satisfaction with life, with extraversion and neuroticism acting as fixed and random effects, clustered by individual. This can be represented by the following equation: Model = lmer (SatisfactionWithLife ~ Extraversion + Neuroticism + (Extraversion + Neuroticism|ID), data.

**Results**

We visualized our model (Figure 3), which indicated that a normally distributed data set, complying with the assumption of normality. When checking for the assumption of linearity, our plot seemed to comply with this. Finally, the model was also compliant with the assumption of homoskedasticity.

We used the package “flexplot” (Fife, 2022) to obtain estimates of the model. According to our model, the estimated values tell us that the variance of individuals’ satisfaction with life is 2.03. This accounts for ~79% of the variance that is explained between reporting of extraversion that is left after accounting for our fixed effects.

The model, as pictured in Table 2 which we obtained using the package “apaTable”, explains roughly 26% (*CI* [.24,.28]) of the variance in satisfaction with life, considering a 95% confidence interval. This suggests that both extraversion and neuroticism have a large effect in predicting life satisfaction. Semi-partial r2 values indicate that neuroticism is responsible for contributing the most to the variance in life satisfaction (*r2* = 0.12, *CI* [.10, .14]), when in comparison to extraversion (*r2* = 0.16, *CI* [.05, .07]).

The data also revealed a significant *b-*weight for both extraversion (*b* = 0.27 *CI* [0.24, .30]) and neuroticism (*b* = - 0.38 *CI* [-0.41, -0.35]), considering a 95% confidence interval. The results indicate extraversion to be significantly positively associated with satisfaction with life. Meanwhile, the inverse pattern is seen for neuroticism. According to our coefficients, we can determine that for every one-unit increase in extraversion, satisfaction with life increases by 0.16. In a mirroring pattern, with every one-unit increase in neuroticism, life satisfaction decreases by -0.21. As described by the previously stated evidence, we can determine, that extraversion is positively associated with satisfaction with life, while neuroticism is negatively associated with it, confirming the previously stated hypothesis 1 & 2.

Residual R2 reports revealed that the model successfully explained 26% of the variance in satisfaction with life within individuals (level 1). Conversely, our model did not meaningfully explain the variability of life satisfaction between individuals (level 2). Therefore, in our model, effects for level 1 were significant, and those for level 2 were not. In line with this, we reject the hypothesis that the effects between level 1 and level 2 will be similar.

Overall, we can conclude that extraversion, positively impacts satisfaction with life, while neuroticism negatively impacts it, and that these effects are meaningful within individuals, but not between them.

**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. [doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).

Fife, D. (2022). Flexplot: Graphically-based data analysis. Psychological Methods, 27(4), 477–496. https://doi.org/10.1037/met0000424

Pedersen T (2024). patchwork: The Composer of Plots. R package version 1.2.0.9000, https://github.com/thomasp85/patchwork, <https://patchwork.data-imaginist.com>.

Stanley D (2021). \_apaTables: Create American Psychological Association (APA) Style Tables\_. R package version 2.0.8, <https://CRAN.R-project.org/package=apaTables>.

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

Table 1

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

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

*Regression results using swl as the criterion.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Predictor | *b* | *b*  95% CI  [LL, UL] | *sr2* | *sr2*  95% CI  [LL, UL] | Fit |
| (Intercept) | 4.62\*\* | [4.43, 4.80] |  |  |  |
| tipm.E | 0.27\*\* | [0.24, 0.30] | .06 | [.05, .07] |  |
| tipm.N | -0.38\*\* | [-0.41, -0.35] | .12 | [.10, .14] |  |
| tipm.N + tipm.E | idTRUE | NA | [NA, NA] | NA | [NA, NA] |  |
|  |  |  |  |  | *R2*  = .263\*\* |
|  |  |  |  |  | 95% CI[.24,.28] |
|  |  |  |  |  |  |

*Note.* A significant *b*-weight indicates the semi-partial correlation is also significant. *b* represents unstandardized regression weights. *sr2* represents the semi-partial correlation squared. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.  
\* indicates p < .05. \*\* indicates p < .01.

Figure 1

*Univariate distribution of variables.*

A graph of numbers and a number of numbers

Description automatically generated with medium confidence

*Note.* “tipm.E” represents extraversion “tipm.N” represents neuroticism, and “swl” represents satisfaction with life.

Figure 2

*Model Summary.*

A graph of different types of lines

Description automatically generated with medium confidence

*Note.* “tipm.E” represents extraversion “tipm.N” represents neuroticism, and “swl” represents satisfaction with life. The above figure represents the model, in which the outcome variable is satisfaction with life, with extraversion and neuroticism acting as fixed and random effects, clustered by individual.

Figure 3

*Key Estimates Summary.*

A group of graphs showing different values

Description automatically generated with medium confidence

*Note.* “tipm.E” represents extraversion “tipm.N” represents neuroticism, and “swl” represents satisfaction with life.