PSYR6003\_Assignment\_4

Alireza Aleali

## **Loading Required Packages**

library(tidyverse)

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(haven)  
library(lme4)

Loading required package: Matrix  
  
Attaching package: 'Matrix'  
  
The following objects are masked from 'package:tidyr':  
  
 expand, pack, unpack

library(flexplot)

Attaching package: 'flexplot'  
  
The following object is masked from 'package:ggplot2':  
  
 flip\_data

library(ggeffects)  
library(patchwork)  
library(e1071)

## **Reading and Preparing Data**

data = read\_sav("/Users/alireza/Desktop/Rclass/Assignment/4th\_assignment/P6003.A4.sav")  
  
#Converts the id column to a factor and Selects only the necessary columns for the analysis, which simplifies the dataset  
  
data = data %>%   
 mutate(id = as.factor(id)) %>%   
 select(id, day, swl, tipm.E, tipm.N)   
  
#Data is already in long format which means no need for converting!

## **Testing Hypotheses and Model Building**

Test a model where extraversion (tipm.E) and neuroticism (tipm.N) are predictors, and the dependent variable is satisfaction with life (swl).

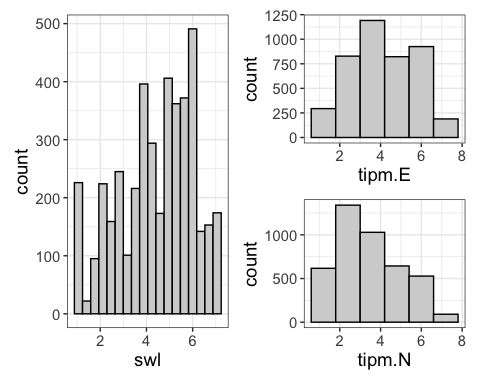
H1. Extraversion will be positively associated with satisfaction with life.

H2. Neuroticism will be negatively associated with satisfaction with life.

H3. The effects will be similar for both level 1 (within participants over time) and level 2 (between participants).

## **Initial Data Exploration with Univariate Plots**

# Using flexplot for visualization  
a = flexplot(swl~1, data = data)  
b = flexplot(tipm.E~1, data = data)  
c = flexplot(tipm.N~1, data = data)  
  
# Arranging plots  
a+b/c



The distribution of satisfaction with life scores may not perfectly follow a normal distribution as it shows a slight skew towards the left. The scores for Extraversion appear to be roughly normal. On the other hand, Neuroticism exhibits a slight skew towards the right. There do not appear to be any significant outliers in the data.

## **Baseline Mixed Model (proof of essential need for mixed effect model)**

# Basic model with only random intercepts for 'id'  
baseline <- lmer(swl ~ 1+(1|id), data = data)  
  
# Calculate Intraclass Correlation Coefficient  
icc\_value=icc(baseline)  
cluster\_sizes <- table(data$id) # Creates a table of the number of occurrences of each cluster ID  
average\_cluster\_size <- mean(cluster\_sizes) # needed for calculating Design Effect  
  
  
# Calculate DEFF using the ICC and average cluster size  
class(icc\_value)

[1] "list"

icc\_value <- as.numeric(icc\_value) #since it's not numeric we need to convert it!  
deff\_calculated <- 1 + (average\_cluster\_size - 1) \* icc\_value  
  
print(paste("Average Cluster Size:", average\_cluster\_size))

[1] "Average Cluster Size: 16.1673003802281"

print(paste("ICC:", icc\_value))

[1] "ICC: 0.741986261766127" "ICC: 12.2968824358209"

print(paste("Calculated Design Effect:", deff\_calculated))

[1] "Calculated Design Effect: 12.2539285102094"  
[2] "Calculated Design Effect: 187.510509644447"

The analysis of the baseline mixed model revealed an **Intraclass Correlation Coefficient (ICC) of 0.74**, indicating that approximately 74% of the variance in life satisfaction scores is due to differences between subjects. This high ICC value underscores substantial similarities within each subject compared to variations between different subjects. Additionally, a **design effect of 12.3** suggests that ignoring the data’s clustered nature would significantly underestimate variance, implying the necessity of using a linear mixed model (LMM). Not using an LMM could lead to pseudoreplication, where the effective sample size is falsely inflated by about 12 times, potentially skewing the statistical significance of the results. Using a **linear mixed model** is crucial for accurately accounting for clustering in the data, ensuring valid statistical analysis and reliable study conclusions.

## **Exploratory Model Building (Testing Random x Fixed effects in model)**

#add extraversion as fixed effect  
extra\_fixed = lmer(swl ~ tipm.E + (1|id), data = data)  
#add it as a random effect  
extra\_random = lmer(swl ~ tipm.E + (tipm.E|id), data = data)   
#comparing the models  
model.comparison(extra\_fixed, extra\_random)

refitting model(s) with ML (instead of REML)

$statistics  
 aic bic bayes.factor p  
extra\_fixed 10945.67 10971.08 0.000000e+00 <2e-16  
extra\_random 10865.86 10903.99 3.720188e+14   
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.009 0.033 0.087 1.254   
  
$r\_squared\_change  
(Intercept) Residual   
 0.29714234 -0.06066128

#same process with neuroticism, to see model fit #consider neuroticism as a fixed effect  
neuro\_fixed = lmer(swl ~ tipm.E + tipm.N + (tipm.E|id), data = data)

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
Model failed to converge with max|grad| = 0.00358453 (tol = 0.002, component 1)

model.comparison(extra\_random, neuro\_fixed)

refitting model(s) with ML (instead of REML)

$statistics  
 aic bic bayes.factor p  
extra\_random 10865.86 10903.99 0.00000e+00 <2e-16  
neuro\_fixed 10498.33 10542.81 2.67951e+78   
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.055 0.125 0.231 1.035   
  
$r\_squared\_change  
(Intercept) tipm.E Residual   
 0.1503284 0.1793129 0.0797324

#Considering neuroticism as a random effect  
neuro\_random = lmer(swl ~ tipm.E + tipm.N + (tipm.E+tipm.N|id), data = data)

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
Model failed to converge with max|grad| = 0.00351972 (tol = 0.002, component 1)

model.comparison(neuro\_fixed, neuro\_random)

refitting model(s) with ML (instead of REML)

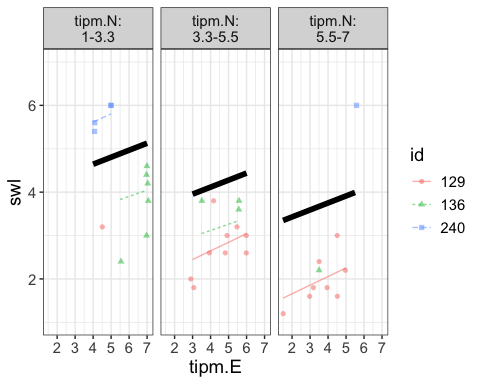
$statistics  
 aic bic bayes.factor p  
neuro\_fixed 10498.33 10542.81 0.000000e+00 <2e-16  
neuro\_random 10364.47 10428.02 8.457116e+24   
  
$predicted\_differences  
 0% 25% 50% 75% 100%   
0.000 0.011 0.038 0.110 1.195   
  
$r\_squared\_change  
 (Intercept) tipm.E Residual   
 0.007567003 -0.281457650 -0.080672336

#Final model  
model = lmer(swl ~ tipm.E + tipm.N + (tipm.E+tipm.N|id), data = data)

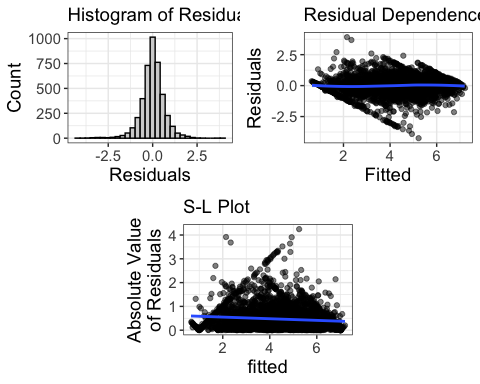
Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
Model failed to converge with max|grad| = 0.00351972 (tol = 0.002, component 1)

Upon evaluating models, it was determined that treating extraversion as a random effect yielded better results, evidenced by lower AIC and BIC values, a higher Bayes factor, and significant p-values. Further analysis included adding neuroticism to the model, which significantly improved the model’s performance based on the same metrics. Consequently, the final model that treats both extraversion and neuroticism as random effects was deemed most effective.

visualize(model, plot = "model")



visualize(model, plot = "residuals" )



#checking kurtosis  
res = residuals(model)  
kurtosis(res)

[1] 4.153064

The histogram of the residuals displays a bell-shaped distribution but exhibits leptokurtosis (the kurtosis is greater than 3). The plot of residuals against predicted values shows a slight pattern, which is generally undesirable, though the zero slope suggests that the data are sufficiently linear. The scale-location plot reveals a minimal slope, indicating little heteroscedasticity. Independence is no longer considered an assumption in this context.

summary(model)

Linear mixed model fit by REML ['lmerMod']  
Formula: swl ~ tipm.E + tipm.N + (tipm.E + tipm.N | id)  
 Data: data  
  
REML criterion at convergence: 10344.5  
  
Scaled residuals:   
 Min 1Q Median 3Q Max   
-6.0752 -0.4530 0.0362 0.4921 5.6013   
  
Random effects:  
 Groups Name Variance Std.Dev. Corr   
 id (Intercept) 2.02500 1.4230   
 tipm.E 0.01983 0.1408 -0.41   
 tipm.N 0.03510 0.1874 -0.32 -0.05  
 Residual 0.48874 0.6991   
Number of obs: 4249, groups: id, 262  
  
Fixed effects:  
 Estimate Std. Error t value  
(Intercept) 4.50550 0.11825 38.10  
tipm.E 0.16138 0.01530 10.55  
tipm.N -0.21052 0.01708 -12.32  
  
Correlation of Fixed Effects:  
 (Intr) tipm.E  
tipm.E -0.621   
tipm.N -0.518 0.131  
optimizer (nloptwrap) convergence code: 0 (OK)  
Model failed to converge with max|grad| = 0.00351972 (tol = 0.002, component 1)

The summary output indicates that the coefficients for Extraversion and Neuroticism exceed their respective standard errors, suggesting that both have statistically significant slopes.

Variability attributed to random effects:

totalVar = 2.02500 + 0.01983 + 0.03510 + 0.48874  
#Total variance = 2.56867  
(idVar = (2.02500/totalVar)\*100)

[1] 78.83457

a substantial portion of the variation in the data, specifically 78.83%, is attributed to differences among participants, as indicated by the participant ID treated as a random effect. After adjusting for fixed effects in the model, participant ID still accounts for 79% of the unexplained variance, highlighting its significant influence on the outcomes of the model.

#### **Model estimation**

estimates(model)

Note: your models were fit to two different datasets.  
This is \*probably\* because you have missing data in one, but not the other.  
I'm going to make the dangerous assumption this is the case and do some ninja moves  
 in the background (hiya!). If you don't want me to do this, handle the missing data in advance

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
Model failed to converge with max|grad| = 0.00351972 (tol = 0.002, component 1)

refitting model(s) with ML (instead of REML)

Fixed Effects:   
(Intercept) tipm.E tipm.N   
 4.5055041 0.1613842 -0.2105195   
  
  
Random Effects:   
 Groups Name Std.Dev. Corr   
 id (Intercept) 1.42302   
 tipm.E 0.14082 -0.407   
 tipm.N 0.18735 -0.324 -0.052  
 Residual 0.69910   
  
  
ICC and Design Effect:   
 icc design.effect   
 0.7419863 12.2968824   
  
  
R Squared:   
  
(Intercept) Residual   
-0.06062994 0.26049350

library(performance)

Attaching package: 'performance'

The following object is masked from 'package:flexplot':  
  
 icc

r2(model)

# R2 for Mixed Models  
  
 Conditional R2: 0.788  
 Marginal R2: 0.094

For each one-unit increase in extraversion, life satisfaction increases by 0.16, accounting for neuroticism. Conversely, each one-unit increase in neuroticism decreases life satisfaction by 0.21, accounting for extraversion.

The R-squared residual (within individuals) is 26%, indicating that the model accounts for 26% of the variance in life satisfaction from one time point to another within the same individual.

The R-squared intercept (across individuals) is 0%, suggesting that the model does not explain any of the variance in life satisfaction between different individuals. Extraversion and neuroticism do not contribute to understanding differences in life satisfaction across individuals.

However, the model does a reasonable job of explaining variability in life satisfaction within an individual over time, capturing 26% of this variability. Extraversion and neuroticism are significant in explaining these within-individual differences over time.

The Conditional R-squared, which represents the proportion of variance explained by both fixed and random effects, is 0.788, indicating a substantial overall model fit. The Marginal R-squared, representing the variance explained by fixed effects alone, is 0.094, showing a more modest contribution from the fixed effects.

#### **Summary Statistics**

library(apaTables)  
  
dataReduced = data %>%   
 drop\_na() %>%  
 select(id, swl, tipm.E, tipm.N)   
head(dataReduced)

# A tibble: 6 × 4  
 id swl tipm.E tipm.N  
 <fct> <dbl> <dbl> <dbl>  
1 15 2.6 2 6.5  
2 15 1.6 3.5 7   
3 15 2.2 2 6   
4 15 3 4 5.5  
5 15 5.2 5.5 3.5  
6 15 2.6 6 4

table1 = apa.cor.table(dataReduced, table.number=1, filename = "Correlation\_Table")  
print(table1)

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.  
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.

table1 = as.tibble(table1$table.body)

Warning: `as.tibble()` was deprecated in tibble 2.0.0.  
ℹ Please use `as\_tibble()` instead.  
ℹ The signature and semantics have changed, see `?as\_tibble`.

table1 = table1 %>%   
 mutate(row = c(1:8)) %>%   
 filter(row == "1" |row == "3" |row == "4"|row == "6"|row == "7") %>%   
 select(!(row))  
print(table1)

# A tibble: 5 × 5  
 Variable M SD `1` `2`   
 <chr> <chr> <chr> <chr> <chr>   
1 "1. swl" "4.43" "1.61" " " " "   
2 "2. tipm.E" "4.18" "1.52" ".38\*\*" " "   
3 " " " " " " "[.35, .41]" " "   
4 "3. tipm.N" "3.49" "1.54" "-.45\*\*" "-.33\*\*"   
5 " " " " " " "[-.47, -.43]" "[-.36, -.31]"

write.csv(table1, file="table1.csv")

#### **Results of the Model**

library(broom.mixed)  
  
tidyModel = tidy(model, effects = "fixed", conf.int=TRUE)  
print(tidyModel)

# A tibble: 3 × 7  
 effect term estimate std.error statistic conf.low conf.high  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
1 fixed (Intercept) 4.51 0.118 38.1 4.27 4.74   
2 fixed tipm.E 0.161 0.0153 10.5 0.131 0.191  
3 fixed tipm.N -0.211 0.0171 -12.3 -0.244 -0.177

est = estimates(model)

Note: your models were fit to two different datasets.  
This is \*probably\* because you have missing data in one, but not the other.  
I'm going to make the dangerous assumption this is the case and do some ninja moves  
 in the background (hiya!). If you don't want me to do this, handle the missing data in advance

Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
Model failed to converge with max|grad| = 0.00351972 (tol = 0.002, component 1)

refitting model(s) with ML (instead of REML)

icc = as\_tibble(est$icc[1], rownames="icc")  
rsq = as\_tibble(est$r.squared, rownames = "R squared")  
  
  
Table2 = dplyr::bind\_rows(tidyModel, icc, rsq)  
print(Table2)

# A tibble: 6 × 10  
 effect term estimate std.error statistic conf.low conf.high icc value  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <dbl>  
1 fixed (Interce… 4.51 0.118 38.1 4.27 4.74 <NA> NA   
2 fixed tipm.E 0.161 0.0153 10.5 0.131 0.191 <NA> NA   
3 fixed tipm.N -0.211 0.0171 -12.3 -0.244 -0.177 <NA> NA   
4 <NA> <NA> NA NA NA NA NA icc 0.742   
5 <NA> <NA> NA NA NA NA NA <NA> -0.0606  
6 <NA> <NA> NA NA NA NA NA <NA> 0.260   
# ℹ 1 more variable: `R squared` <chr>

write.csv(Table2, file="table2.csv")