

# Daily Diary - Longitudinal Data Analysis

*Isidro Landa*

*September 28, 2017*

## Load packages

```
rm(list=ls())

library(haven)
library(tidyverse)
library(magrittr)
library(ggplot2)
library(lme4)
library(psych)
library(egghelper)
library(knitr)
library(kableExtra)
```

```
## Warning: package 'kableExtra' was built under R version 3.4.2
```

## Data

```
## Set working directory
setwd("C:/Users/Isidro/OneDrive/WUSTL/Coursework/F17/ALDA/Assignments/1-descriptives-and-graphs-isidrol")
#setwd("Z:/Lab/Lab Members/Isidro/Assignments/")

erdata <- read_sav("Landa-CC daily diary.sav")
post <- read_sav("Campus Culture-Post malone.sav")

erdata <- merge(erdata, post, all.x = T, by = "id")
# Merge time-invariant variables from post-survey
erdata %<>% arrange(id, day)
erdata[is.na(erdata)] <- NA
# Imported data had NaNs, but only is.na worked, not is.nan
```

## Framework & Study variables

The current analyses invoke a Dual Process Model of Homesickness (DPM-HS; Stroebe, Schut, & Nauta, 2016) as a framework. The aim of these analyses are 1) to better understand effect of rumination on homesickness and 2) to determine what kind of new place adjustment (i.e., academic or social) is most relevant for this association.

The DPM-HS suggests that there are home-factors (e.g., thoughts about missing people back home) that are at the essence of homesickness and new place factors (e.g., adjustment) that should moderate the homesickness.

```
text_tbl <- data.frame(
  Variable = c("Day (day)", "Homesick (e.homesick)", "Rumination (er.rumin)",
    "Academic Functioning (acadfx)", "Belonging Uncertainty (bun_m)"),
  Items = c(
    "Monday - Friday",
    "To what extent did you experience the following emotions today? - Homesick",
    "How much did you do any of the following in order to manage how you felt today? - I thought over and over about the situation and my feelings",
    "Composite: 1) How lost did you feel in your assignments?, 2) How much did you procrastinate on your assignments?, 3) How uncertain did you feel of where your academic future/career was headed?, 4) How satisfied were you with your academic performance?",
    "Composite: 1) Sometimes I feel that I belong at Wash U, and sometimes I feel that I don't belong at Wash U, 2) When something bad happens, I feel that maybe I don't belong at Wash U"),
  Scale = c("1 - 7", "1 (Not at all) - 5 (A great deal)",
    "1 (Not at all) - 5 (A great deal)", "1 (Not at all) - 5 (Very Much)",
    "1 (Strongly disagree) - 7 (Strongly agree)")
)

kable(text_tbl, format = "latex", booktabs = T) %>%
kable_styling(full_width = F, latex_options = c("striped", "scale_down")) %>%
column_spec(1, bold = T) %>%
column_spec(2, width = "30em")
```

Variable	Items	Scale
Day (day)	Monday - Friday	1 - 7
Homesick (e.homesick)	To what extent did you experience the following emotions today? - Homesick	1 (Not at all) - 5 (A great deal)
Rumination (er.rumin)	How much did you do any of the following in order to manage how you felt today? - I thought over and over about the situation and my feelings	1 (Not at all) - 5 (A great deal)
Academic Functioning (acadfx)	Composite: 1) How lost did you feel in your assignments?, 2) How much did you procrastinate on your assignments?, 3) How uncertain did you feel of where your academic future/career was headed?, 4) How satisfied were you with your academic performance?	1 (Not at all) - 5 (Very Much)
Belonging Uncertainty (bun_m)	Composite: 1) Sometimes I feel that I belong at Wash U, and sometimes I feel that I don't belong at Wash U, 2) When something bad happens, I feel that maybe I don't belong at Wash U	1 (Strongly disagree) - 7 (Strongly agree)

## Insert rows for missing time points (if needed)

This makes it possible to test lagged effects

```
# CC_new_rows <- tbl_df(expand.grid(unique(day$id), seq(1, 7, 1)) %>%  
#   rename(id = Var1, day = Var2) %>%  
#   full_join(day))  
# # tbl_df (or can use 'as_tibble') creates a new data frame  
#  
# erdata <- arrange(CC_new_rows, id, day)
```

## Graphs

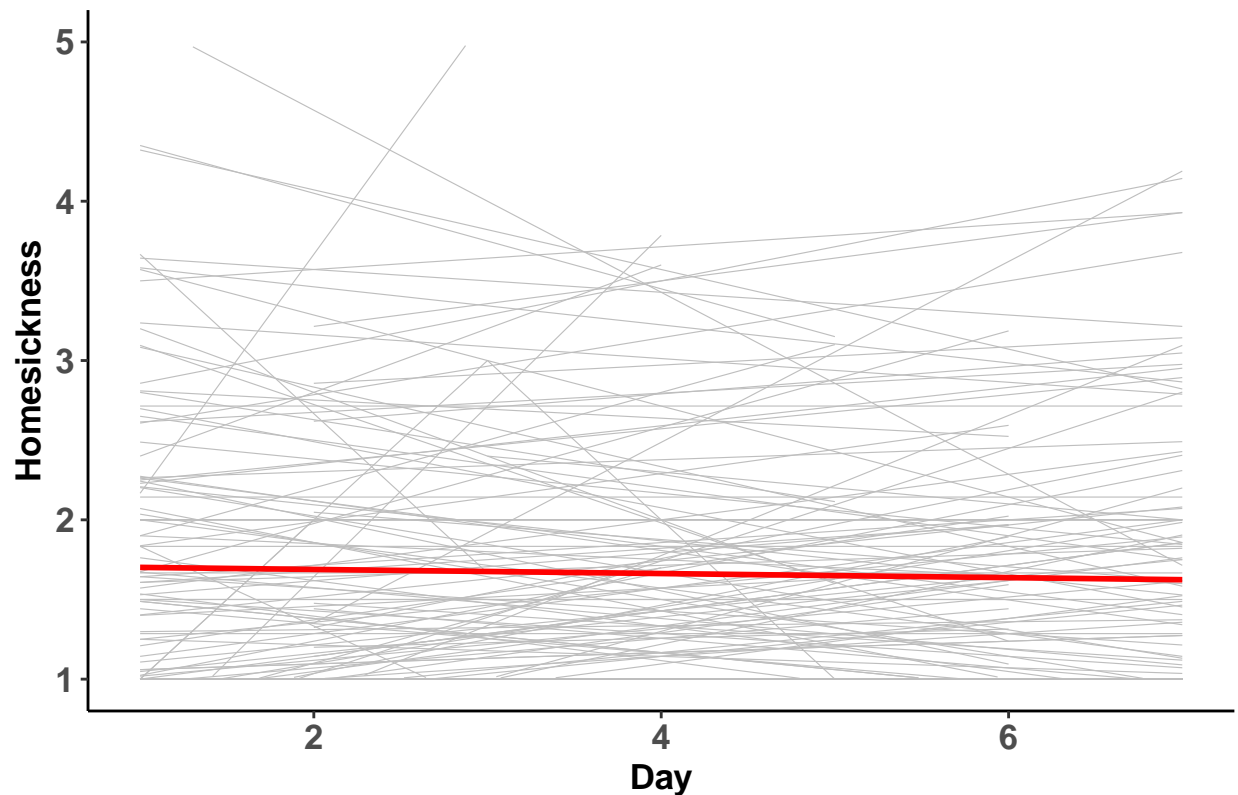
```
# Spaghetti plot with average growth trajectory  
erdata %>%  
  ggplot(aes(x = day, y = e.homesick)) +  
    geom_smooth(aes(group = id), method = "lm", se = F, color = "gray", size = .2) +  
    geom_smooth(method = "lm", se = F, color = "red", size = 1) +  
    labs(x = "Day", y = "Homesickness", title = "Figure 2. Growth Curve") +  
    ylim(1, 5) +  
    theme_classic() +  
    theme(legend.position = "none",  
          axis.text = element_text(face = "bold", size = rel(1.2)),  
          axis.title = element_text(face = "bold", size = rel(1.2)),  
          plot.title = element_text(face = "bold", size = rel(1.2), hjust = .06))
```

```
## Warning: Removed 244 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 244 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 876 rows containing missing values (geom_smooth).
```

**Figure 2. Growth Curve**



```
# Different color for each participant

# erdata %>%
#   ggplot(erdata, aes(day, er.rumin, group = id)) +
#   geom_point() + stat_smooth(method = "lm", se = F) +
#   aes(color = factor(id)) + guides(color = F) +
#   # Coloured by participant
#   theme_classic() +
#   xlab("Day") + ylab("Emotion Regulation-Rumination") + ylim(1, 5) +
#   ggtitle("Rumination as a Function of Day")

# Generate a random sample because 150 too many for graph below
set.seed(24)
er_24 <- erdata %>%
  select(id) %>%
  distinct %>%
  sample_n(24)

erdata2 <- left_join(er_24, erdata)

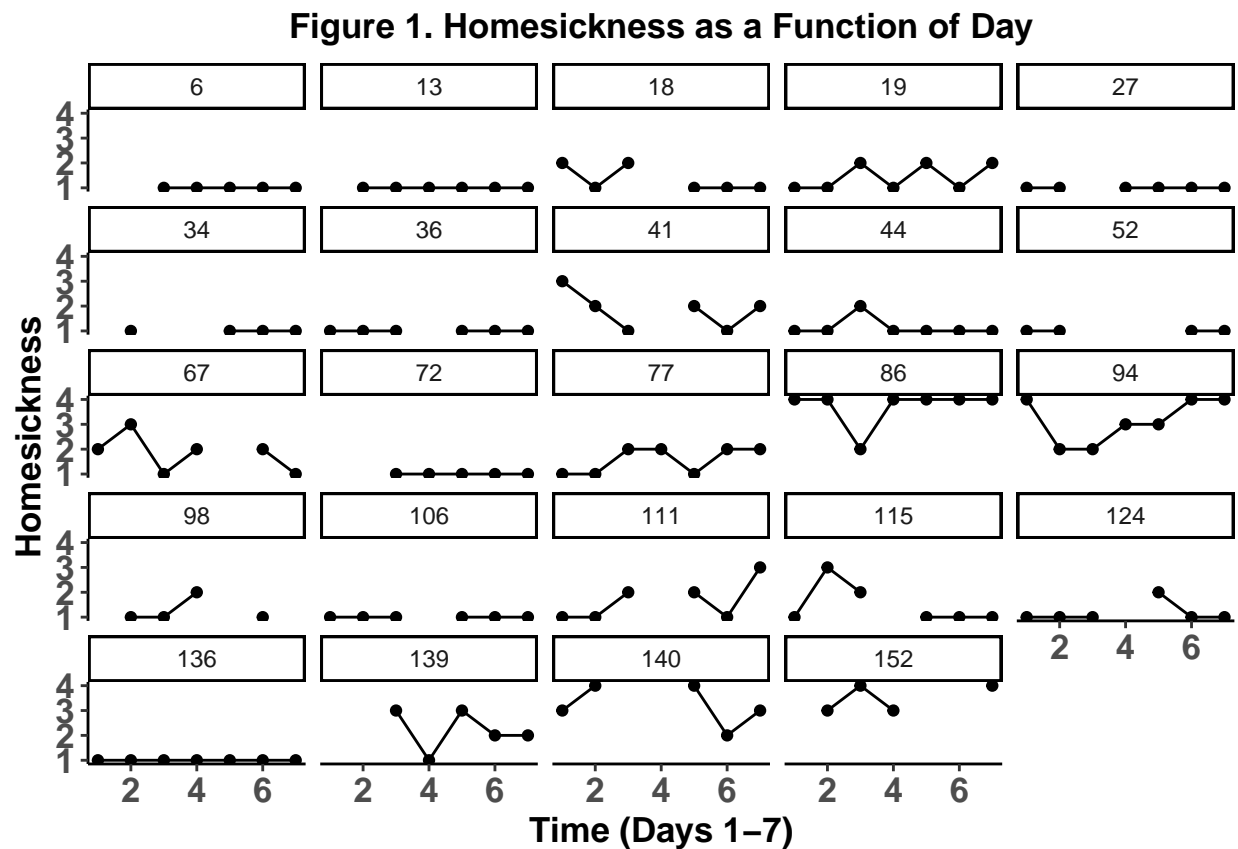
## Joining, by = "id"
# Use left_join to keep all observations for each time point
# only for ids that were sampled

# Daily fluctuation in homesickness
```

```
ggplot(erdata2, aes(day, e.homesick, group = id)) + geom_line() +
  geom_point() + facet_wrap(~id) +
  xlab("Time (Days 1-7)") + ylab("Homesickness") +
  ggtitle("Figure 1. Homesickness as a Function of Day") +
  theme_classic() +
  theme(legend.position = "none",
        axis.text = element_text(face = "bold", size = rel(1.2)),
        axis.title = element_text(face = "bold", size = rel(1.2)),
        plot.title = element_text(face = "bold", size = rel(1.2),
                                   hjust = .5))
```

## Warning: Removed 11 rows containing missing values (geom\_path).

## Warning: Removed 30 rows containing missing values (geom\_point).



## Compute variables

```
# Belonging uncertainty (BUN) and academic functioning (acad.fx) composites
erdata %<>% mutate(bun_m = bun1 + bun2 / 2,
                  acadfx = acad.lost + acad.procrast + acad.uncertain +
                    acad.perform.r / 4)
```

```
# Correlations/internal consistency among items
```

```
# acad_df <- erdata %>% select(acad.procrast:acad.perform)
# keys = c(1, 1, 1, -1)
# psych::alpha(x = acad_df, keys = keys)
# alpha = 0.68
# bun_df <- erdata %>% select(bun1, bun3)
# correlate(bun_df)
# r = 0.69
```

## Rescaling of predictors

```
# Make time interpretable making day 1 zero
# erdata$day0 <- erdata$day - 1

# Calculate aggregates within a separate dataframe
# i.e., Mean for each person across their respective time points
erdata_agg <- erdata %>% group_by(id) %>%
  summarize(er.rumin_agg = mean(er.rumin, na.rm = T),
    e.homesick_agg = mean(e.homesick, na.rm = T))

# Merge with original dataset
erdata <- merge(erdata, erdata_agg, all.x = T, by = "id")

# Person- and grand mean-center using original and aggregated scores
erdata %<>% mutate(er.rumin_pc = er.rumin - er.rumin_agg,
  e.homesick_pc = e.homesick - e.homesick_agg,
  bun_gm = bun_m - mean(bun_m, na.rm = T),
  acadfx_gm = acadfx - mean(acadfx, na.rm = T))

# Subtract person's mean from their raw score to get person-centered
# For bun_gm & acadfx: grand mean center bC they're time-invariant

# Check calculations; should get 'TRUE' if done correctly
# all(erdata$er.rumin == erdata$er.rumin_pc + erdata$er.rumin_agg,
#   na.rm = T)
```

## Concurrent correlations & separating sources of variance

```
# Unconditional means model (i.e., intercept and time only)
mod1a <- lmer(e.homesick ~ day0 + (1 | id), erdata)
summary(mod1a)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + (1 | id)
## Data: erdata
##
## REML criterion at convergence: 1813.2
##
## Scaled residuals:
## Min 1Q Median 3Q Max
```

```
## -3.8682 -0.4406 -0.1360  0.4895  4.1028
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   id       (Intercept) 0.4653    0.6822
##   Residual                0.3781    0.6149
## Number of obs: 806, groups:  id, 150
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  1.656912   0.067872  24.412
## day0         0.002755   0.011064   0.249
##
## Correlation of Fixed Effects:
##      (Intr)
## day0 -0.467
```

*# Intercept: Mean of homesickness is 1.66 when day = 0 (i.e., day 1)*  
*# Day: NS; but would be interpreted as an increase in homesickness each day*

*ICC(mod1a) # Plenty of within-person variance but slightly more bw-person*

```
## [1] 0.551699
```

*# Concurrent correlation between rumination and homesickness*  
mod1b <- lmer(e.homesick ~ day0 + er.rumin + (1 | id), erdata)  
summary(mod1b)

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + er.rumin + (1 | id)
##      Data: erdata
##
## REML criterion at convergence: 1800.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8846 -0.4634 -0.1350  0.4752  3.8153
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   id       (Intercept) 0.4429    0.6655
##   Residual                0.3789    0.6155
## Number of obs: 800, groups:  id, 150
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  1.463659   0.090561  16.162
## day0         0.005018   0.011125   0.451
## er.rumin     0.082527   0.026526   3.111
##
## Correlation of Fixed Effects:
##              (Intr) day0
## day0        -0.381
## er.rumin    -0.675  0.045
```

```

# Intercept: Mean of homesickness is 1.46 at day 1,
# controlling for rumination
# Rumination: Difficult to interpret because an uncentered variable
# conflates two sources of variance (between & within)

# Now including aggregated (between) person-centered (within) rumination
mod1c <- lmer(e.homesick ~ day0 + er.rumin_agg + er.rumin_pc + (1 | id), erdata)
summary(mod1c)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + er.rumin_agg + er.rumin_pc + (1 | id)
## Data: erdata
##
## REML criterion at convergence: 1799.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.8880 -0.4672 -0.1257  0.5105  3.8661
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## id      (Intercept)  0.4312    0.6567
## Residual                    0.3787    0.6154
## Number of obs: 800, groups: id, 150
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.134368   0.176529   6.426
## day0         0.004858   0.011121   0.437
## er.rumin_agg 0.229142   0.072571   3.157
## er.rumin_pc  0.060543   0.028437   2.129
##
## Correlation of Fixed Effects:
##              (Intr) day0   er.rmn_g
## day0         -0.192
## er.rumin_gg  -0.927  0.013
## er.rumin_pc  -0.008  0.043  0.001

```

```

# Intercept: Mean of homesickness is 1.13 at day 1,
# controlling for rumination
# Rumination_agg: 0.22 increase in homesickness for 1-unit increase in rumination
# Rumination_pc: 0.06 CONCURRENT association with homesickness
# On days that rumination goes up, homesickness does too (not by much after bw)

```

## Examining predictions from week to week

For example, does Rumination on day one predict Homesickness on day two, controlling for Homesickness on day one, and so forth.

```

erdata %<>% mutate(home_prev = lag(e.homesick),
                  rumin_prev = lag(er.rumin),
                  homepc_prev = lag(e.homesick_pc),

```



```

        ruminpc_prev = lag(er.rumin_pc)) %>%
  select(er.rumin, rumin_prev, id, day, e.homesick, home_prev, everything())
erdata$ruminpc_prev[erdata$day == 1] <- NA
erdata$homepc_prev[erdata$day == 1] <- NA

```

## Time-lagged models

```

# Does rumination predict homesickness controlling for previous day homesickness?
mod2a <- lmer(e.homesick ~ day0 + homepc_prev + ruminpc_prev + (1 | id), erdata)
summary(mod2a)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + homepc_prev + ruminpc_prev + (1 | id)
## Data: erdata
##
## REML criterion at convergence: 1267.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7516 -0.3911 -0.1419  0.3450  3.5676
##
## Random effects:
## Groups Name Variance Std.Dev.
## id      (Intercept) 0.4813  0.6938
## Residual              0.3508  0.5923
## Number of obs: 555, groups: id, 148
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.6231340  0.0803673  20.196
## day0         0.0009704  0.0153341   0.063
## homepc_prev -0.0807683  0.0475416  -1.699
## ruminpc_prev 0.0536311  0.0350419   1.530
##
## Correlation of Fixed Effects:
##              (Intr) day0  hmcp_p
## day0         -0.615
## homepc_prev -0.011  0.075
## ruminpc_prev -0.033  0.055 -0.032

```

*# No*

```

# Only when you account for individual differences in growth curves
# is it that you get lagged effects
mod2b <- lmer(e.homesick ~ day0 + homepc_prev + ruminpc_prev + (1 + day0 | id), erdata)
summary(mod2b)

```

```

## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + homepc_prev + ruminpc_prev + (1 + day0 |
## id)
## Data: erdata
##

```

```

## REML criterion at convergence: 1245.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5790 -0.3689 -0.1233  0.3843  3.0133
##
## Random effects:
##   Groups   Name                Variance Std.Dev. Corr
##   id       (Intercept) 0.83694  0.9148
##           day0         0.02197  0.1482  -0.66
##   Residual                0.27666  0.5260
## Number of obs: 555, groups: id, 148
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.6172830  0.0924855  17.487
## day0         -0.0003926  0.0193788  -0.020
## homepc_prev  -0.2014276  0.0463399  -4.347
## ruminpc_prev  0.0507869  0.0338507   1.500
##
## Correlation of Fixed Effects:
##              (Intr) day0   hmcp_p
## day0         -0.734
## homepc_prev  -0.002  0.046
## ruminpc_prev -0.027  0.041 -0.023

# Homesickness: Negative autocorrelation?
# Positive (0.49) when not pc

# Does this change with the introduction of a time-invariant moderator?
mod2c <- lmer(e.homesick ~ day0 + homepc_prev + ruminpc_prev*bun_gm +
              (1 | id), erdata)
summary(mod2c)

## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + homepc_prev + ruminpc_prev * bun_gm + (1 |
##      id)
##      Data: erdata
##
## REML criterion at convergence: 1134.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7899 -0.3779 -0.1302  0.3354  3.6188
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   id       (Intercept) 0.4164  0.6453
##   Residual                0.3431  0.5857
## Number of obs: 504, groups: id, 131
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)  1.559826  0.081641  19.106
## day0         0.003645  0.015812   0.231

```

```

## homepc_prev          -0.057969   0.049711  -1.166
## ruminpc_prev         0.039357   0.037014   1.063
## bun_gm               0.056263   0.034162   1.647
## ruminpc_prev:bun_gm  0.011782   0.019071   0.618
##
## Correlation of Fixed Effects:
##           (Intr) day0   hmcp_p rmnpc_ bun_gm
## day0      -0.634
## homepc_prev -0.015  0.082
## ruminpc_prev -0.039  0.063 -0.040
## bun_gm      0.000  0.010 -0.006  0.006
## rmnpc_prev:_ -0.018  0.038  0.082 -0.108 -0.004

# Nope; No cross-level interaction

# What about academic adjustment?
mod2d <- lmer(e.homesick ~ day0 + homepc_prev + ruminpc_prev*acadfx_gm +
              (1 | id), erdata)
summary(mod2d)

## Linear mixed model fit by REML ['lmerMod']
## Formula: e.homesick ~ day0 + homepc_prev + ruminpc_prev * acadfx_gm +
##          (1 | id)
## Data: erdata
##
## REML criterion at convergence: 1127.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.7777 -0.4063 -0.1472  0.3294  3.5638
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  id       (Intercept)  0.4158     0.6448
##  Residual                    0.3462     0.5883
## Number of obs: 499, groups: id, 129
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    1.563106   0.081963  19.071
## day0           0.004739   0.015937   0.297
## homepc_prev    -0.060996   0.049789  -1.225
## ruminpc_prev    0.039954   0.038023   1.051
## acadfx_gm       0.043370   0.023653   1.834
## ruminpc_prev:acadfx_gm 0.005833   0.014197   0.411
##
## Correlation of Fixed Effects:
##           (Intr) day0   hmcp_p rmnpc_ acadfx_
## day0      -0.630
## homepc_prev -0.013  0.078
## ruminpc_prev -0.036  0.053 -0.024
## acadfx_gm   -0.015  0.013 -0.009  0.019
## rmnpc_prev:_ -0.018  0.057 -0.030 -0.233  0.025

```

*# No cross-level interaction here either*

### Potential conclusions:

Perhaps it's a combination of having slightly more between-person variance and the time-scale (daily vs weekly). Homesickness seems to have more variability on a weekly time-scale ( $ICC = .43$ ). Despite ML estimation, it could also be the small sample size. There were  $N = 150$  for daily portion and  $N = 135$  for analyses including time-invariant variables. Also, rumination seems to have a large between effect.