STATISTICAL HORIZONS

Intensive Longitudinal Methods

Donald Hedeker, Ph.D.

Upcoming Seminar:
November 15-16, 2019, Philadelphia, Pennsylvania

Intensive Longitudinal Methods: Mixed Models for Ecological Momentary Assessment (EMA) Data

Donald Hedeker
Department of Public Health Sciences
University of Chicago

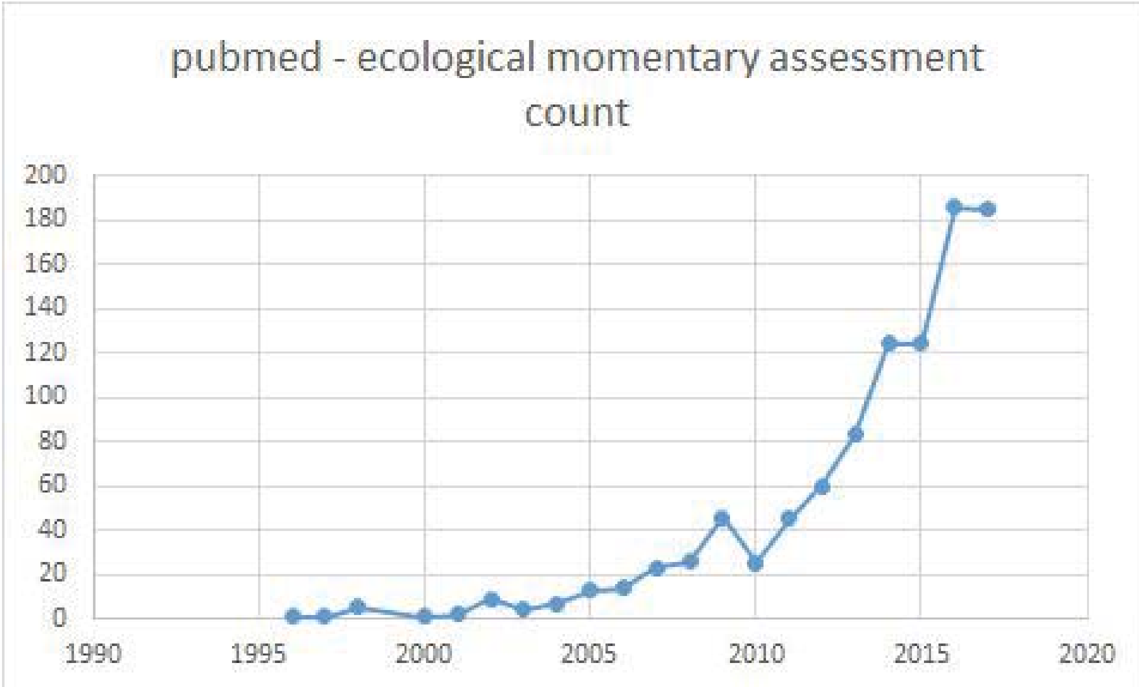
https://hedeker-sites.uchicago.edu

Supported by NHLBI grant R01 HL121330 (Hedeker & Dunton).

Ecological Momentary Assessment (EMA) data

aka experience sampling and diary methods

- Subjects provide frequent reports on events and experiences of their daily lives (*e.g.*, 30-40 responses per subject collected over the course of a week or so)
- electronic diaries: palm pilots, personal digital assistants (PDAs), interactive voice response (IVR) systems, cell phones, actigraphs, web-based
- Capture particulars of experience in a way not possible with more traditional designs
  e.g., allow investigation of phenomena as they happen over time
- Reports could be time-based, following a fixed-schedule, randomly triggered, event-triggered



Per year counts of articles from a pubmed search

Data are rich and offer many modeling possibilities!

- person- and occasion-level effects on occasion-level responses
  ⇒ potential influence of context and/or environment
  e.g., subject response might vary when alone vs with others
- data are inherently multilevel
  - occasions (level-1) within subjects (level-2)
  - occasions (level-1) within days (level-2) within subjects (level-3)
  - occasions (level-1) within waves (level-2) within subjects (level-3)
- References for mixed model analysis of EMA data
  - Schwartz, J.E. & Stone, A. (2007). The analysis of real-time momentary data: A practical guide. In: A.A. Stone, S.S. Shiffman, A. Atienza, and L. Nebeling, editors, *The science of real-time data capture: Self-report in health research*. Oxford, England: Oxford University Press, p. 76-113.
  - Walls, T.A., Jung, H., & Schwartz, J.E. (2006). Multilevel models for intensive longitudinal data. In: Walls, T.A. and Schafer, J.L., editors, *Models for intensive longitudinal data*. New York: Oxford University Press, p. 3-37

Mixed model applications

- To produce summary statistics, accounting for clustering of data
- Examine within-subject effect of time-varying covariates
- Examine why subjects differ in mean level as well as variability
  - between-subjects variance
    e.g., subject heterogeneity could vary by gender or age
  - within-subjects variance
    e.g., subject degree of stability could vary by gender or age
- To examine intercept and slope heterogeneity in terms of covariates
- Modeling of the timing of event reports
  - Time until first event in a day
  - Event times during and across days

Ecological Momentary Assessment (EMA) Study of Adolescent Smokers (Mermelstein)

- 461 students completed the baseline measurement wave
- participants were in either 9th or 10th grade at baseline
- 55.1% female
- reported on a screening questionnaire 6-8 weeks prior to baseline that they had smoked at least one cigarette in their lifetime
- 57.6% smoked at least one cigarette in the past month at baseline
- written parental consent and student assent were required
- 57% were white, 20% hispanic, 16% black, and 7% of other race

Random prompts and Smoking events

- Carry PDA for a week, answer questions when prompted

  average = 30 answered prompts (median = 30, range = 7 to 71)

  $\Sigma_i^N n_i = 14,105$ total number of random prompts

- Subjects trained to to event record smoking episodes

  234 subjects provided at least one smoking event
  average = 4.9 smoking events (median = 3, range = 1 to 42)

  $\Sigma_i^N n_i = 1,142$ total number of smoking events

- Mutually exclusive

- $N = 234$, Spearman corr of n(random) with n(smoke) = -.08 (ns)

Dependent Variables - Random prompt versions

- Positive Affect mood scale (mean=6.797 and sd=1.935)
  - Before signal: I felt Happy
  - Before signal: I felt Relaxed
  - Before signal: I felt Cheerful
  - Before signal: I felt Confident
  - Before signal: I felt Accepted by Others

- Negative Affect mood scale (mean=3.455 and sd=2.253)
  - Before signal: I felt Sad
  - Before signal: I felt Stressed
  - Before signal: I felt Angry
  - Before signal: I felt Frustrated
  - Before signal: I felt Irritable

⇒ items rated on 1 (not al all) to 10 (very much) scale

For smoking events, subjects rated mood before and after smoking

Summary statistic for the outcomes

y_{ij} = affect level of subject i at occasion j

which mean do you report?

- $\bar{y}_{..}$ = mean ignoring clustering of data
- mean of $\bar{y}_{i.}$ = mean of subject-level means

are they different?

suppose that there are two subjects with values

- 1
- 10, 10, 10, 10, 10, 10, 10, 10

- $\bar{y}_{..} = 9$
- mean of $\bar{y}_{i.} = 5.5$

$\Rightarrow$ Equal only if the number of obs is the same for all subjects (which usually doesn't happen with EMA)

which to use?

Neither!

Want some kind of estimate that takes into account

- clustering of data within subjects and correlation of the clustered observations
- varying numbers of observations per subject

Mixed model ($i = 1, 2, \ldots N$ subjects, and $j = 1, 2, \ldots n_i$ obs)

$$y_{ij} = \beta_0 + \upsilon_i + \epsilon_{ij} \qquad \text{where } \upsilon_i \sim N(0, \sigma^2_\upsilon) \text{ and } \epsilon_i \sim N(0, \sigma^2_\epsilon)$$

$\hat{\beta}_0$ is such an estimate

$\Rightarrow$ intercept in a mixed model with no covariates

$$\hat{\beta}_0 = \frac{\bar{y}_{..} - \frac{1}{N^*} \Sigma_{i=1}^{N} \, n_i \, \rho_i \, \bar{y}_{i.}}{1 - \frac{1}{N^*} \Sigma_{i=1}^{N} \, n_i \, \rho_i}$$

where

- $\rho_i = n_i r / [1 + (n_i - 1) r]$ = Spearman-Brown reliability
- r = intraclass correlation $\quad r = \frac{\sigma^2_\upsilon}{\sigma^2_\upsilon + \sigma^2_\epsilon}$
- N^* = total number of observations $= \Sigma_{i=1}^{N} n_i$

note,
$\hat{\beta}_0 = \bar{y}_{..}$ if $n_i = n$ or $r = \rho_i = 0$

SAS MIXED syntax

```
PROC MIXED COVTEST;
CLASS id;
MODEL NegAff = / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;
```

`id` = subject id variable
`NegAff` = negative affect (occasion-varying) score

⇒ mean, accounting for clustering, is intercept estimate

Does it matter?

Random prompts ($N = 461, \Sigma n_i = 14105$)

	NA	PA
$\bar{y}_{..}$	3.455	6.797
Avg($\bar{y}_{i.}$)	3.485	6.777
mixed model $\hat{\beta}_0$	3.483	6.779

Smoking events, mood rating BEFORE smoking
($N = 234, \Sigma n_i = 1141$)

	NA	PA
$\bar{y}_{..}$	3.493	6.604
Avg($\bar{y}_{i.}$)	3.984	6.384
mixed model $\hat{\beta}_0$	3.908	6.411

SAS example: MoodMeans.sas

```
DATA one; INFILE 'U:\Data\Robin\PreMoods\MoodRanPreSmk.dat';
INPUT id PosAff NegAff SmkE;
PROC FORMAT;
     VALUE SmkE 0 = 'Random' 1 = 'Smoke';

PROC SORT; BY SmkE;

/* raw mood means for random and smoking events */
PROC MEANS; VAR PosAff NegAff;
BY SmkE; FORMAT Smke Smke.;
RUN;
```



```
---------------------------------------- SmkE=Random ----------------------------------------

                              The MEANS Procedure

      Variable         N          Mean       Std Dev       Minimum       Maximum

      PosAff       14105     6.7969656     1.9346636     1.0000000    10.0000000
      NegAff       14105     3.4548458     2.2528845     1.0000000    10.0000000


---------------------------------------- SmkE=Smoke -----------------------------------------

      Variable         N          Mean       Std Dev       Minimum       Maximum

      PosAff        1141     6.6036810     2.1001806     1.0000000    10.0000000
      NegAff        1141     3.4934268     2.3633460     1.0000000    10.0000000
```

```
/* means of subject mood means for random and smoking events */
PROC SORT; BY SmkE id;
PROC MEANS NOPRINT; BY SmkE id; VAR PosAff NegAff;
OUTPUT OUT = summdat MEAN(PosAff NegAff) = MPosAff MNegAff;
PROC MEANS DATA=summdat; VAR MPosAff MNegAff;
BY SmkE; FORMAT Smke Smke.;
RUN;
```



```
---------------------------------------- SmkE=Random ----------------------------------------

                              The MEANS Procedure

        Variable       N            Mean         Std Dev         Minimum         Maximum

        MPosAff      461       6.7774679       1.2355936       2.3882353       9.8160000
        MNegAff      461       3.4849921       1.5255682       1.0190476       8.1833333


---------------------------------------- SmkE=Smoke -----------------------------------------

        Variable       N            Mean         Std Dev         Minimum         Maximum

        MPosAff      234       6.3840328       1.7822418       1.5000000      10.0000000
        MNegAff      234       3.9839033       2.1026653       1.0000000       9.8000000
```

```
/* mixed model estimates of means - random prompts */
DATA random; SET one; IF SmkE=0;
PROC MIXED COVTEST;
CLASS id;
MODEL PosAff =  / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;

PROC MIXED COVTEST;
CLASS id;
MODEL NegAff =  / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;
RUN;

/* mixed model estimates of means - smoking events */
DATA smoke; SET one; IF SmkE=1;
PROC MIXED COVTEST;
CLASS id;
MODEL PosAff =  / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;

PROC MIXED COVTEST;
CLASS id;
MODEL NegAff =  / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;
RUN;
```


The Mixed Procedure

Dimensions

Subjects	461
Max Obs Per Subject	71

Number of Observations

Number of Observations Used	14105

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr > Z
Intercept	id	1.4451	0.1006	14.36	<.0001
Residual		2.3028	0.02788	82.60	<.0001

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	6.7793	0.05756	460	117.78	<.0001

Stata example: MoodMeans.do

```
cd "U:\Data\Robin\PreMoods"
log using MoodMeans.log, replace
infile id PosAff NegAff SmkE using MoodRanPreSmk.dat, clear

label define SmkElabel 0 "random prompt" 1 "smoke event"
label values SmkE SmkElabel

* get subject mood means for random and smoke
sort id SmkE
egen MPosAff = mean(PosAff), by(id SmkE)
egen MNegAff = mean(NegAff), by(id SmkE)

* get one mean per subject for random and smoke
bys id SmkE: replace MPosAff = . if _n>1
bys id SmkE: replace MNegAff = . if _n>1

local myvars "PosAff NegAff MPosAff MNegAff"
* raw mood means and mean of subject mood means for random and smoke
sort SmkE
by SmkE: summ `myvars'
```



```
-> SmkE = random prompt

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      PosAff |     14,105    6.796966    1.934664          1         10
      NegAff |     14,105    3.454846    2.252885          1         10
     MPosAff |        461    6.777468    1.235594   2.388235      9.816
     MNegAff |        461    3.484992    1.525568   1.019048   8.183333

-------------------------------------------------------------------------------
-> SmkE = smoke event

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      PosAff |      1,141    6.603681    2.100181          1         10
      NegAff |      1,141    3.493427    2.363346          1         10
     MPosAff |        234    6.384033    1.782242        1.5         10
     MNegAff |        234    3.983903    2.102665          1        9.8
```

```
* mixed model estimates of means - random prompts
mixed PosAff if SmkE==0  || id:
mixed NegAff if SmkE==0  || id:
* mixed model estimates of means - smoking events
mixed PosAff if SmkE==1  || id:
mixed NegAff if SmkE==1  || id:

log close
```



```
Mixed-effects ML regression                     Number of obs     =     14,105
Group variable: id                              Number of groups  =        461

                                                Obs per group:
                                                              min =          7
                                                              avg =       30.6
                                                              max =         71

                                                Wald chi2(0)      =          .
Log likelihood = -26580.659                     Prob > chi2       =          .

------------------------------------------------------------------------------
      PosAff |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   6.779296   .0575026   117.90   0.000     6.666593    6.891999
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
id: Identity                 |
                  var(_cons) |   1.442183   .1003463      1.258329    1.652899
-----------------------------+------------------------------------------------
               var(Residual) |   2.302813   .0278801      2.248813    2.358111
------------------------------------------------------------------------------
LR test vs. linear model: chibar2(01) = 5482.66       Prob >= chibar2 = 0.0000
```

Why are the smoking means so different?

NA: $\bar{y}_{..} = 3.493$ $\quad$ $\text{Avg}(\bar{y}_{i.}) = 3.984$ $\quad$ $\hat{\beta}_0 = 3.908$ mixed model
PA: $\bar{y}_{..} = 6.604$ $\quad$ $\text{Avg}(\bar{y}_{i.}) = 6.384$ $\quad$ $\hat{\beta}_0 = 6.411$ mixed model

$\Rightarrow$ subjects with many smoking events have lower NA and higher PA

Consider $\texttt{n\_smk}_i$ (# of smoking events) as a regressor in a model of mood before smoking:

$$\texttt{Mood}_{ij} = \beta_0 + \beta_1 \texttt{n\_smk}_i + v_i + \epsilon_{ij}$$

NA: $\hat{\beta}_1 = -.0661$, $se = .019$, $z = -3.48$, $p < .0005$
PA: $\hat{\beta}_1 = .0311$, $se = .017$, $z = 1.87$, $p < .062$

SAS example: `MoodMeans.sas`

```
/* get number of smoking events per subject */
PROC SORT DATA=one; BY id;
PROC MEANS NOPRINT; BY id; VAR SmkE;
   OUTPUT OUT = nummdat    SUM(SmkE) = NSmkE;
/* Select smoking events for subjects with smoking events */
DATA two; MERGE one nummdat; BY id;
IF NSmkE > 0 AND SmkE=1;
/* mixed model analysis of NegAff with NSmkE */
PROC MIXED COVTEST;
CLASS id;
MODEL NegAff = NSmkE / SOLUTION;
RANDOM INTERCEPT / SUBJECT=id;
RUN;
```

Stata example:`MoodMeans.do`

```
* mixed model analysis of NegAff with NSmkE
sort id
egen NSmkE = sum(SmkE), by (id)
mixed NegAff NSmkE if SmkE==1  || id:
```

Example for exercises (Mermelstein et al, 2002)

- 8th or 10th graders carried hand-held computers during a seven consecutive day data collection period
- 17,514 random prompts from 515 students; average of 34 prompts per student (range = 3 to 58).
- Outcome is measure of the subject's positive mood (`posmood` mean=6.733, sd=2.117)
- Of interest is the degree of heterogeneity in positive mood in both WS and BS variation
- Covariates: `genderf` coded 0 for males and 1 for females; `alone` a prompt-varying covariate coded 0 if the subject was alone or 1 if with others

posmood_example.dat (`id, posmood, alone, genderf`)

posmood_example.dat - Notepad

File Edit Format View Help

```
 1  9.6667  0  1
 1 10.0000  1  1
 1  9.0000  0  1
 1  7.3333  0  1
 1  8.6667  0  1
 1 10.0000  0  1
 1  7.6667  1  1
 1  5.6667  0  1
 1  8.0000  0  1
 1  6.6667  1  1
 1  8.3333  1  1
 1  7.6667  1  1
 1  4.0000  1  1
 1  8.0000  1  1
 1  5.0000  1  1
 1  9.3333  0  1
 1  8.3333  0  1
 1  7.6667  1  1
 1  6.3333  1  1
 1  7.3333  0  1
 1  8.3333  1  1
 1  7.0000  0  1
 1  7.3333  1  1
 1  6.6667  0  1
 1  6.6667  1  1
 1  4.3333  0  1
 2 10.0000  0  1
 2  9.3333  0  1
 2  9.0000  0  1
 2  3.6667  0  1
 2  9.6667  0  1
 2 10.0000  0  1
 2 10.0000  1  1
 2 10.0000  1  1
 2 10.0000  0  1
 2 10.0000  0  1
 2  7.0000  0  1
 2 10.0000  0  1
 2 10.0000  0  1
 2 10.0000  0  1
 2  9.3333  0  1
 2 10.0000  0  1
 2 10.0000  1  1
 2  7.0000  0  1
```

Exercise

Using `posmood_example.dat` and either SAS or Stata

- Obtain raw means of posmood for with others and alone prompts
- Obtain mean of subject means of posmood for with others and alone prompts
- Obtain mixed model estimates of posmood means for with others and alone prompts

Do they differ?

need help?
syntax examples: `MoodMeans_posmood_example.sas`
and `MoodMeans_posmood_example.do`

Time-varying Covariates - WS and BS effects

Section 4.5.2 in Hedeker & Gibbons (2006), Longitudinal Data Analysis, Wiley.