14. Multilevel Missing Data

1

Multilevel Data

A unit of analysis is the what or whom being studied

Observations, individuals, classrooms, dyads, etc.

Multilevel data structures have multiple units of analysis that are hierarchically nested

Lower-level units are nested in higher-level units

2

Examples

Students nested within classrooms

Clients nested within therapists

Twins nested within dyads

Individuals nested within families

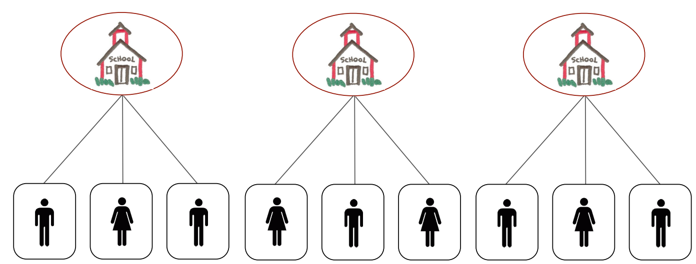
Employees nested within workgroups

Repeated measures nested within individuals

3

Example

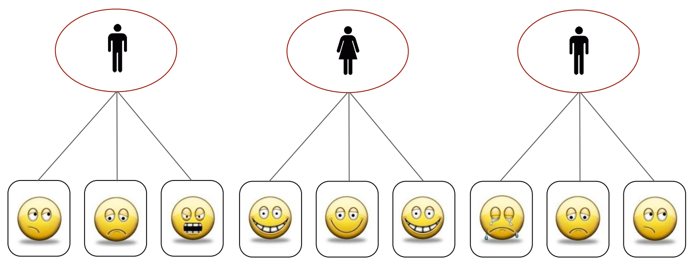
Sample comprised of multiple schools and several students in each school (i.e., students nested within schools)



4

Example

Sample comprised of multiple individuals, each with several daily assessments of mood (i.e., observations nested within individuals)

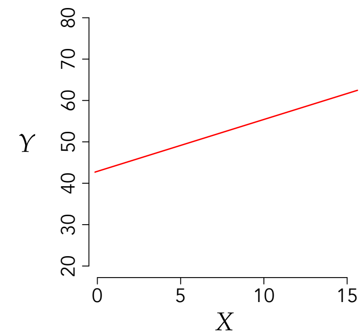


Single-Level Regression

$$Y_i = \beta_0 + \beta_1(X_i) + \varepsilon_i$$

Single-level regression features a common model for all participants

Any pair of observations or scores are independent



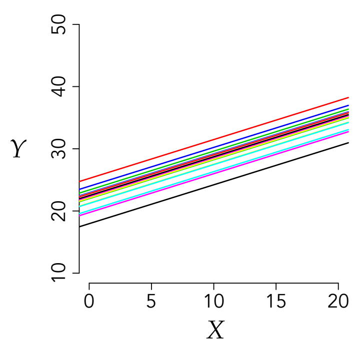
Random Intercepts

$$Y_{ij} = \beta_{0j} + \beta_1\left(X_{ij}\right) + \varepsilon_{ij}$$

Observations are nested in groups

Group regression lines differ in level (the Y mean) but not slope

The influence of the covariate is constant



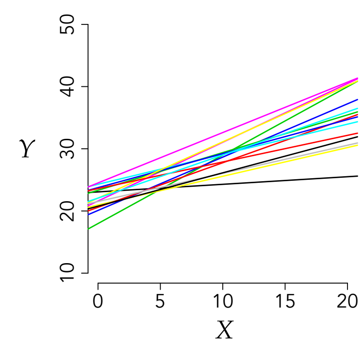
Random Coefficients

$$Y_{ij} = \beta_{0j} + \beta_{1j}\left(X_{ij}\right) + \varepsilon_{ij}$$

Observations are nested in groups

Group regression lines differ in level and slope

The influence of the covariate varies



Missing Data Handling Options

Maximum likelihood works well when missing values are relegated to the outcome variable

Missing predictors can introduce bias, particularly those with random coefficients

Bayesian analyses or multiple imputation is currently a more flexible option

Problem Solving Data

Cluster-randomized study (schools are randomly assigned to intervention or control conditions) of math problem-solving scores

Up to seven repeated measurements nested in 982 students, and students grouped in 29 schools

The data include problem-solving scores and academic-related variables such as math self-efficacy, standardized reading math, and socio-demographic variables

problemsolving.dat

Variable	Name	Missing %	Scaling
School identifier variable	SCHOOL	0	Integer index
Student identifier variable	STUDENT	0	0 = female, 1 = male
Data collection wave	WAVE	0	Integer values of 1 to 7
Experimental condition	CONDITION	0	0 = comparison, 1 = experimental
Percent non-English speakers	ESLPCT	0	Continuous
Ethnicity	ETHNIC	9	1 = white, 2 = black, 3 = Hispanic
Gender	MALE	0	0 = female, 1 = male
Free or reduced lunch	FRLUNCH	4.7	0 = none, 1 = assistance
Achievement group	ACHGROUP	2.1	1 = learning disability, 2 = low achieving, 3 = average achieving
Standardized math	STANMATH	7.4	Continuous
Months since start of year	MONTH0	0	6-point ordinal scale
Months until end of year	MONTH7	0	Continuous
Math problem-solving	PROBSOLV	11.4	Continuous
Math self-efficacy	MATHEFF	11.4	Continuous

Substantive Example

Multilevel longitudinal growth model

Each student has up to seven measures of math problem-solving (repeated measures are observations, students are “groups”)

Do students improve in their math problem-solving over time, and do the intervention and control students change at different rates?

Multilevel Growth Models

Change in problem-solving scores is modeled as a linear function of the passage of time

Seven problem-solving assessments were administered in monthly intervals, so change is "clocked" in months

Growth is defined as the average change in problem-solving per each month of the study

The Temporal Predictor

Elapsed time can be expressed relative to the start or end the school year, winter break, etc.

For simplicity, express the passage of time relative to the beginning of the school year

Centering affects the intercept

Person	Wave	Month	Outcome
1	1	0	55
1	2	1	57
1	3	2	58
1	…	…	…
1	7	6	61
2	1	0	42
2	2	1	39
2	3	2	45
2	…	…	…
2	7	6	46
…	…	…	…
N	1	0	47
N	2	1	47
N	3	2	52
N	…	…	…
N	7	6	57

Within-Person (Level-1) Model

Outcome at month t for student i

Months since baseline

$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$

Monthly change for student i

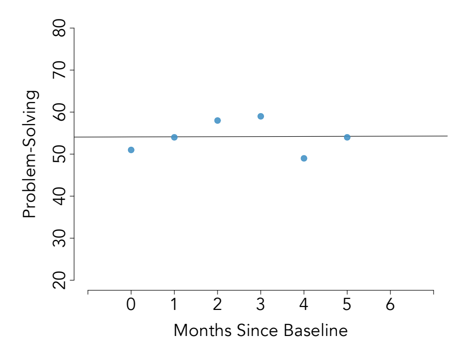
Predicted outcome at month = 0

Residual for student i at month t

Example 1: Data And Linear Trajectory

student	month	probsolv
1	0	51
1	1	54
1	2	58
1	3	59
1	4	49
1	5	54
1	6	NA

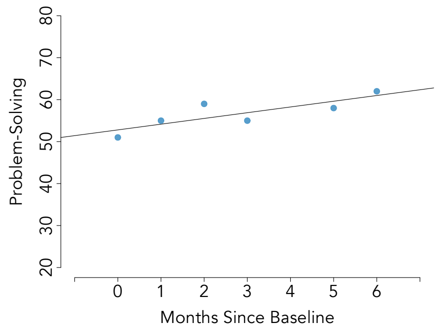
$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 54.10 + .03(MONTH_{ti}) + \varepsilon_{ti}$$



Example 2: Data And Linear Trajectory

student	month	probsolv
5	0	48
5	1	48
5	2	54
5	3	55
5	4	53
5	5	NA
5	6	NA

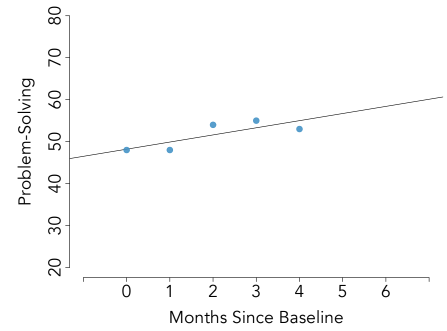
$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 48.20 + 1.70(MONTH_{ti}) + \varepsilon_{ti}$$



Example 3: Data And Linear Trajectory

student	month	probsolv
7	0	52
7	1	54
7	2	61
7	3	52
7	4	67
7	5	44
7	6	45

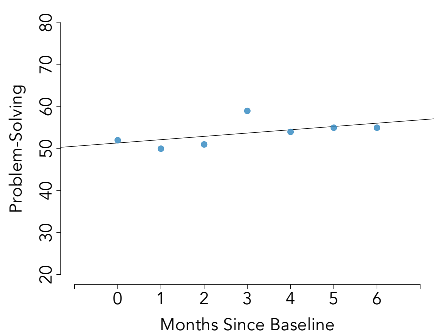
$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 57.32 - 1.25(MONTH_{ti}) + \varepsilon_{ti}$$



Example 4: Data And Linear Trajectory

student	month	probsolv
15	0	48
15	1	47
15	2	45
15	3	51
15	4	49
15	5	47
15	6	56

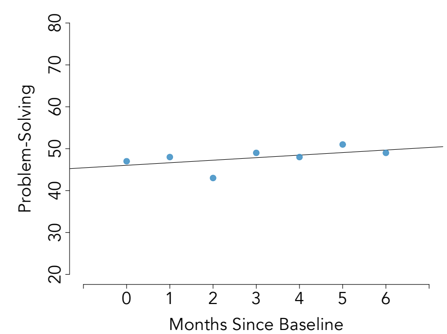
$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 46.00 + 1.00(MONTH_{ti}) + \varepsilon_{ti}$$



Example 5: Data And Linear Trajectory

student	month	probsolv
22	0	43
22	1	49
22	2	47
22	3	46
22	4	41
22	5	51
22	6	38

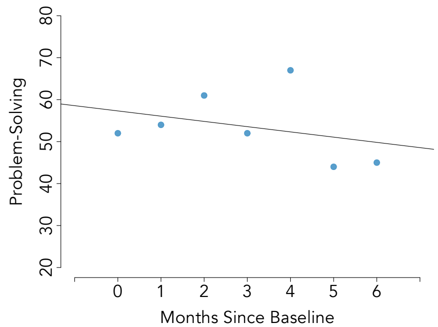
$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 46.82 - .61(MONTH_{ti}) + \varepsilon_{ti}$$



Example 6: Data And Linear Trajectory

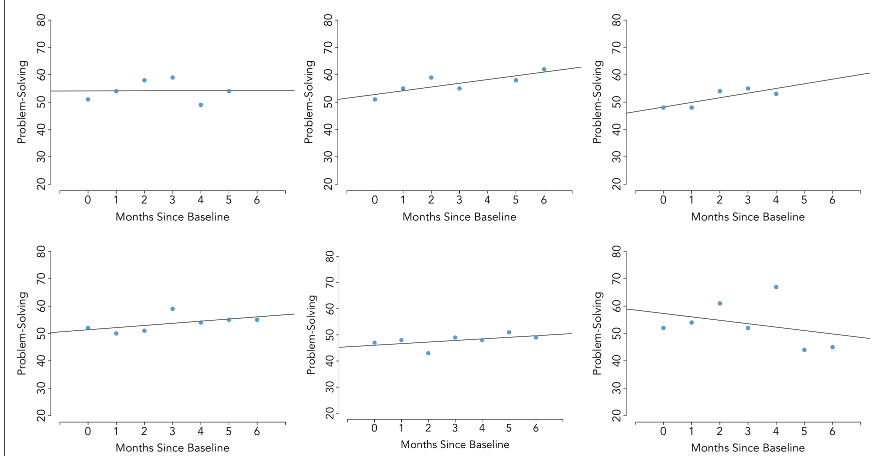
student	month	probsolv
49	0	56
49	1	65
49	2	58
49	3	58
49	4	60
49	5	60
49	6	54

$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$
$$= 60.21 - .50(MONTH_{ti}) + \varepsilon_{ti}$$



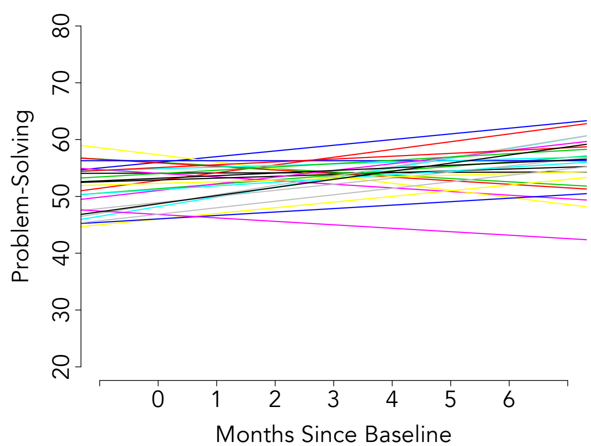
21

Comparison Of Six Trajectories



22

Spaghetti Plot Of 25 Trajectories



23

Between-Person (Level-2) Models

Intercept and slope for student i

Intercept and slope deviations for student i

$$\beta_{0i} = \beta_0 + b_{0i}$$
$$\beta_{1i} = \beta_1 + b_{1i}$$

Mean intercept and slope

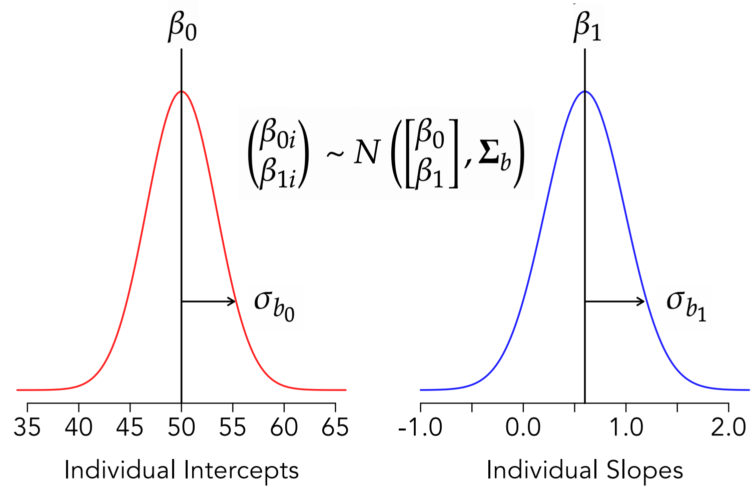
24

Reduced Form Regression Model

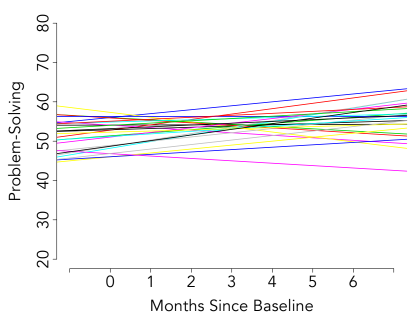
The reduced form model is obtained by replacing the individual intercepts and slopes with their respective between-person models

$$PROBSOLV_{ti} = \underbrace{(\beta_0 + b_{0i})}_{\beta_{0i}} + \underbrace{(\beta_1 + b_{1i})}_{\beta_{1i}}(MONTH_{ti}) + \varepsilon_{ti}$$

Model Summary



Individual Growth Curves Are Imputation Regression Models



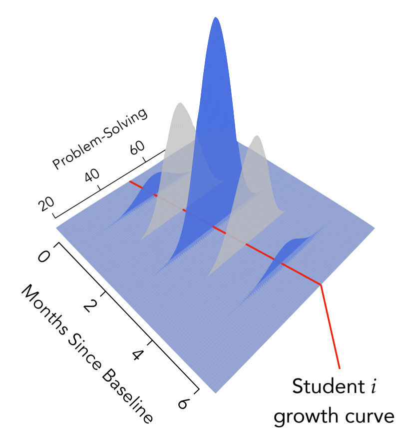
The individual growth curves define regression lines and predicted values for imputation

$$E(Y|X) = \beta_{0i} + \beta_{1i}(MONTH_{ti})$$

Distribution Of Missing Y Scores

Each slice is a normal distribution of Y at a particular assessment for a particular student

Imputations are predicted values + noise, just as before



MCMC Recipe

Do for t = 1 to T iterations

1. Estimate the regression coefficients (average curves), given the filled-in data and other quantities

2. Estimate individual growth curves, given the filled-in data and other quantities

3. Estimate the residual variance, given the filled-in data and other quantities

4. Estimate (impute) missing values, given the regression model parameters

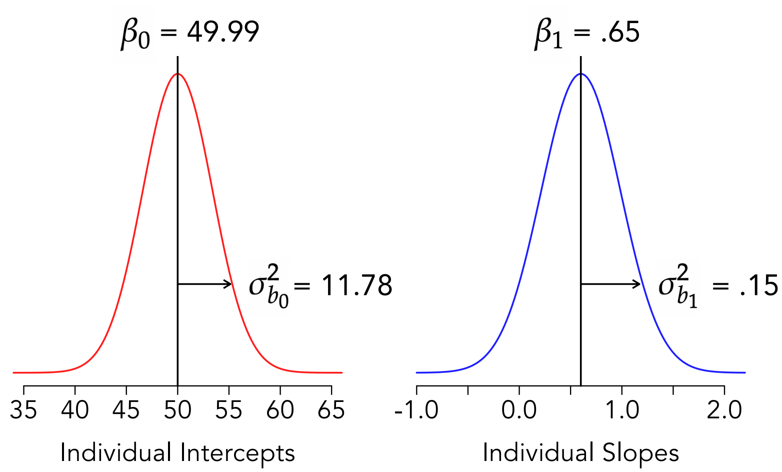
Repeat

Posterior Distribution Summary

Bayesian analysis with 10,000 MCMC iterations

Parameter	Mean	Std. Dev.	Lower 2.5%	Upper 97.5%
Intercept	50.00	0.14	49.73	50.27
MONTH slope	0.65	0.03	0.60	0.71
Intercept variance	11.78	0.84	10.23	13.49
Growth variance	0.15	0.03	0.09	0.22

Visual Summary



Interpretations

The average problem-solving score at the beginning of the school year is β_0 = 50

The variance of the individual intercepts is $\sigma^2_{b_0}$ = 11.78 (standard deviation is $\sqrt{11.78}$ = 3.43)

The average monthly growth rate is β_1 = .65

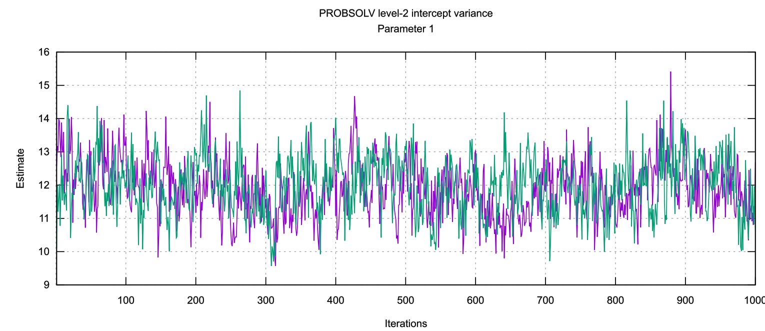
The variance of the individual change rates is $\sigma^2_{b_1}$ = .15 (standard deviation is $\sqrt{.15}$ = .387)

Blimp Bayesian Analysis Script

DATA: problemsolving.dat;
VARIABLES: school student wave condition eslpct ethnic male
frlunch achvgrp stanmath month0 month7 probsolv matheff;
CLUSTERID: student;
FIXED: month0;
MISSING: 999;
MODEL: probsolv ~ month0 | month0;
SEED: 90291;
BURN: 2000;
ITERATIONS: 10000;
CHAINS: 4 processors 4;
OPTIONS: psr;

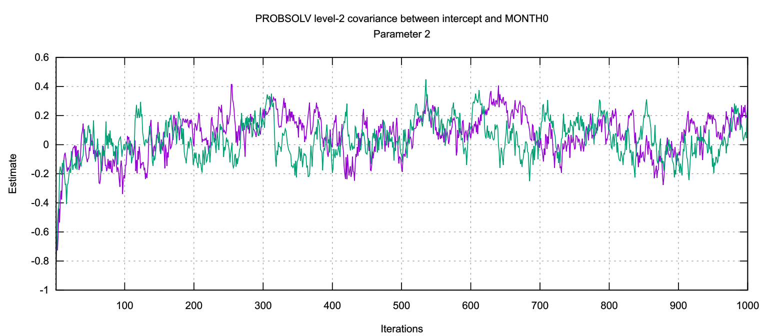
Blimp Trace Plots

Intercept variance estimates from 1000 iterations



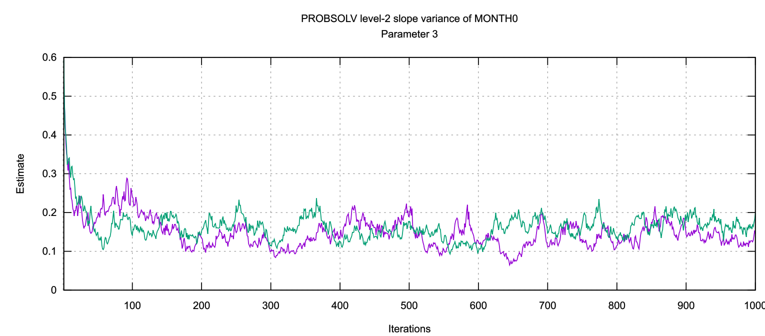
Blimp Trace Plots

Covariance estimates from 1000 iterations



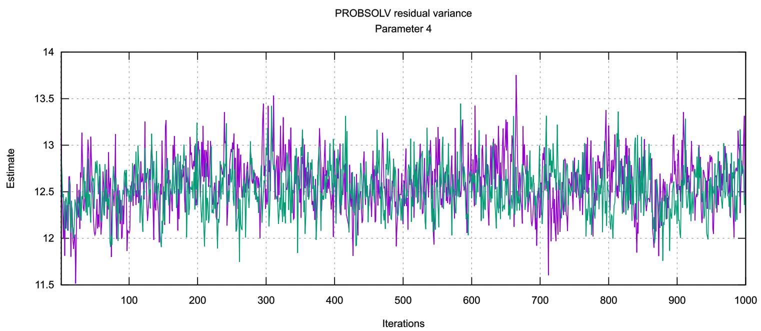
Blimp Trace Plots

Slope variance estimates from 1000 iterations



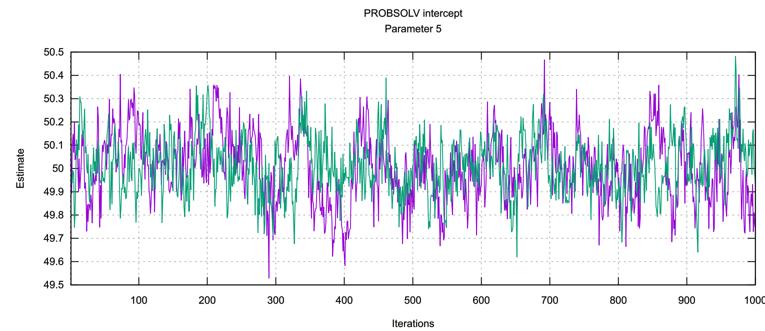
Blimp Trace Plots

Residual variance estimates from 1000 iterations



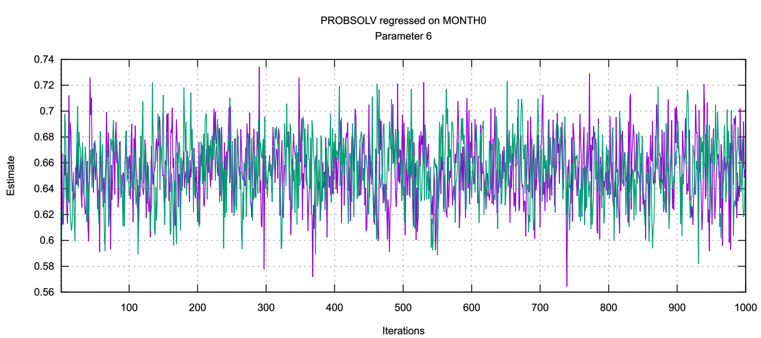
Blimp Trace Plots

Average intercept estimates from 1000 iterations



Blimp Trace Plots

Average slope estimates from 1000 iterations



Blimp Output

```
POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  Comparing iterations across 4 chains   | Highest PSR | Parameter # |
                           51 to 100     |       1.754 |           3 |
                          101 to 200     |       1.016 |           2 |
                          151 to 300     |       1.240 |           3 |
                                ...
                          501 to 1000    |       1.077 |           3 |
                          551 to 1100    |       1.049 |           3 |
                          601 to 1200    |       1.021 |           3 |
                          651 to 1300    |       1.007 |           5 |
                          701 to 1400    |       1.008 |           2 |
                          751 to 1500    |       1.021 |           3 |
                          801 to 1600    |       1.025 |           3 |
                          851 to 1700    |       1.015 |           3 |
                          901 to 1800    |       1.011 |           3 |
                          951 to 1900    |       1.014 |           3 |
                         1001 to 2000    |       1.012 |           3 |
```

Blimp Output

```
ANALYSIS MODEL ESTIMATES:

Missing outcome: probsolv
                                    -----------------------------------------------------------
Parameters                          |   Mean    |  Median   |  StdDev   |Lower 2.5  |Upper 97.5|
                                    -----------------------------------------------------------
Variances:                          |           |           |           |           |          |
  L2 Intercept (i)                  |    11.779 |    11.750 |     0.837 |    10.231 |    13.494|
  L2 (i), month0                    |     0.084 |     0.089 |     0.125 |    -0.174 |     0.312|
  L2 month0                         |     0.149 |     0.148 |     0.032 |     0.092 |     0.216|
  Residual Var.                     |    12.565 |    12.561 |     0.277 |    12.045 |    13.115|
                                    |-----------|-----------|-----------|-----------|----------|
Coefficients:                       |           |           |           |           |          |
  Intercept                         |    49.999 |    49.998 |     0.136 |    49.730 |    50.268|
  month0                            |     0.654 |     0.654 |     0.026 |     0.602 |     0.706|
                                    |-----------|-----------|-----------|-----------|----------|
Standardized Coefficients:          |           |           |           |           |          |
  month0                            |     0.253 |     0.253 |     0.010 |     0.233 |     0.274|
                                    |-----------|-----------|-----------|-----------|----------|
Proportion Variance Explained       |           |           |           |           |          |
  by Fixed Effects                  |     0.064 |     0.064 |     0.005 |     0.054 |     0.075|
  by Level-2 Random Intercepts      |     0.441 |     0.441 |     0.018 |     0.405 |     0.478|
  by Level-2 Random Slopes          |     0.022 |     0.022 |     0.005 |     0.014 |     0.032|
  by Level-1 Residual Variation     |     0.472 |     0.472 |     0.018 |     0.438 |     0.507|
                                    |           |           |           |           |          |
                                    -----------------------------------------------------------
                                    Summaries based on 10000 iterations using 4 chains
```


Group-By-Time (Cross-Level) Interaction

Students were randomly assigned to an intervention or a control (standard) curriculum

The substantive goal is to determine whether the intervention enhances the rates of change

Does the average growth rate differ for intervention and control students?

Within- And Between-Person Models

Within-person growth model

$$PROBSOLV_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$

Between-person model

$$\beta_{0i} = \beta_0 + \beta_2(CONDITION_i) + b_{0i}$$
$$\beta_{1i} = \beta_1 + \beta_3(CONDITION_i) + b_{1i}$$

Reduced-Form Regression Model

$$Y_{ti} = \overbrace{\beta_0 + \beta_1(MONTH_{ti})}^{\text{Expected baseline score and mean monthly growth rate for control group}}$$
$$+ \beta_2(CONDITION_i) \longrightarrow \text{Baseline mean difference}$$
$$+ \beta_3(MONTH_{ti})(CONDITION_i) \longrightarrow \text{Growth rate difference}$$
$$\underbrace{+ b_{0i} + b_{1i}(MONTH_{ti})}_{\text{Individual intercepts and slopes (random effects)}} + \varepsilon_{ti}$$

Blimp Bayesian Analysis Script

```
DATA: problemsolving.dat;
VARIABLES: school student wave condition eslpct ethnic male
  frlunch achvgrp stanmath month0 month7 probsolv matheff;
CLUSTERID: student;
NOMINAL: male frlunch condition;
FIXED: month0 male condition;
MISSING: 999;
MODEL: probsolv ~ month0 condition month0*condition
   male frlunch stanmath | month0;
CENTER: grandmean = male frlunch stanmath;
SIMPLE: month0 | condition;
SEED: 90291;
BURN: 2000;
ITERATIONS: 10000;
CHAINS: 4 processors 4;
OPTIONS: psr;
```


Blimp Output

POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

Comparing iterations across 4 chains	Highest PSR	Parameter #
51 to 100	1.754	3
101 to 200	1.016	2
151 to 300	1.240	3
…		
501 to 1000	1.077	3
551 to 1100	1.049	3
601 to 1200	1.021	3
651 to 1300	1.007	5
701 to 1400	1.008	2
751 to 1500	1.021	3
801 to 1600	1.025	3
851 to 1700	1.015	3
901 to 1800	1.011	3
951 to 1900	1.014	3
1001 to 2000	1.012	3

Blimp Output

ANALYSIS MODEL ESTIMATES:

Missing outcome: probsolv

Grand Mean Centered: frlunch#1 stanmath male

Parameters	Mean	Median	StdDev	Lower 2.5	Upper 97.5
Variances:					
L2 Intercept (i)	5.322	5.312	0.542	4.295	6.415
L2 (i), month0	0.008	0.005	0.105	-0.194	0.212
L2 month0	0.121	0.122	0.032	0.055	0.182
Residual Var.	12.571	12.564	0.277	12.040	13.125
Coefficients:					
Intercept	50.202	50.203	0.198	49.808	50.589
month0	0.434	0.434	0.043	0.350	0.518
condition#1	-0.294	-0.292	0.230	-0.747	0.153
male	-0.016	-0.016	0.184	-0.379	0.340
frlunch#1	-0.141	-0.142	0.248	-0.635	0.351
stanmath	0.026	0.026	0.001	0.024	0.028
month0*condition#1	0.348	0.348	0.054	0.243	0.455
Proportion Variance Explained					
by Fixed Effects	0.325	0.325	0.015	0.296	0.352
by Level-2 Random Intercepts	0.195	0.195	0.017	0.163	0.229
by Level-2 Random Slopes	0.018	0.018	0.005	0.008	0.026
by Level-1 Residual Variation	0.462	0.462	0.016	0.432	0.494

Summaries based on 10000 iterations using 4 chains

Blimp Output

CONDITIONAL EFFECTS ANALYSIS:

Missing outcome: probsolv

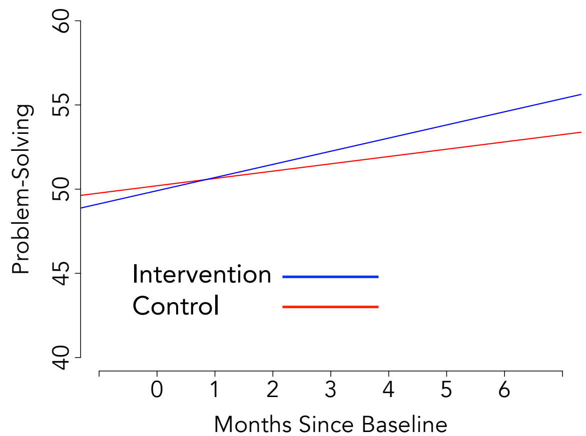
Grand Mean Centered: frlunch#1 stanmath male

Conditional Effects	Mean	Median	StdDev	Lower 2.5	Upper 97.5
month0 | condition#1 @ 0					
Intercept	50.202	50.203	0.198	49.808	50.589
Slope	0.434	0.434	0.043	0.350	0.518
month0 | condition#1 @ 1					
Intercept	49.908	49.909	0.162	49.580	50.217
Slope	0.781	0.781	0.032	0.720	0.843

Summaries based on 10000 iterations using 4 chains

NOTE: Intercepts are computed by setting all predictors not involved in the conditional effect to zero.

Simple Slopes



Posterior Distribution Summary

Bayesian analysis with 10,000 MCMC iterations

Parameter	Mean	Std. Dev.	Lower 2.5%	Upper 97.5%
Intercept	50.20	0.20	49.81	50.59
MONTH slope	0.43	0.04	0.35	0.52
CONDITION slope	-0.29	0.23	-0.75	0.15
MONTH x CONDITION	0.35	0.05	0.24	0.46

Interpretations

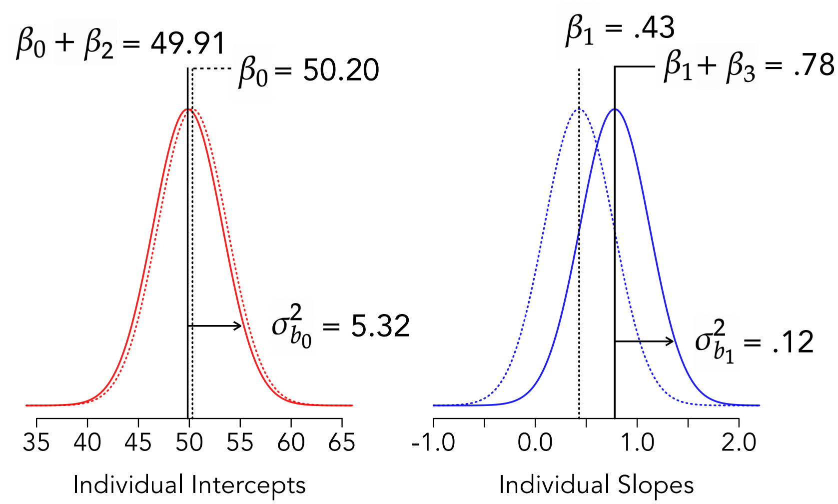
The mean problem-solving score at the beginning of the year for students in the control condition is $\beta_0 = 50.20$

The initial status mean difference for students in the intervention condition is $\beta_2 = -.29$ lower (not significant)

The average monthly growth rate for students in the control condition is $\beta_1 = .43$

The monthly growth rate for students in the intervention condition is $\beta_3 = .35$ higher, on average

Visual Summary



Blimp Model-Based Imputation Script For Analysis In Mplus

```
DATA: problemsolving.dat;
VARIABLES: school student wave condition eslpct ethnic male
  frlunch achvgrp stanmath month0 month7 probsolv matheff;
CLUSTERID: student;
NOMINAL: male frlunch condition;
FIXED: month0 male condition;
MISSING: 999;
MODEL: probsolv ~ month0 condition month0*condition male frlunch stanmath | month0;
SEED: 90291;
NIMPS: 20;
BURN: 2000;
THIN: 2000;
CHAINS: 4 processors 4;
OPTIONS: psr;
SAVE: separate = imps_*.dat;
```


Blimp Model-Based Imputation Script For Analysis in R, SAS, SPSS, Stata

```
DATA: problemsolving.dat;
VARIABLES: school student wave condition eslpct ethnic male
  frlunch achvgrp stanmath month0 month7 probsolv matheff;
CLUSTERID: student;
NOMINAL: male frlunch condition;
FIXED: month0 male condition;
MISSING: 999;
MODEL: probsolv ~ month0 condition month0*condition male frlunch stanmath | month0;
SEED: 90291;
NIMPS: 20;
BURN: 2000;
THIN: 2000;
CHAINS: 4 processors 4;
OPTIONS: psr;
SAVE: stacked0 = imps_stacked.dat;
```


Output Data Information

Stacked file format (R, SAS, SPSS, Stata)

```
VARIABLE ORDER IN SAVED DATA:

   imp# school student wave condition eslpct ethnic male frlunch achgroup
     stanmath month0 month7 probsolv matheff;
```

Separate file format (Mplus)

```
VARIABLE ORDER IN SAVED DATA:

   school student wave condition eslpct ethnic male frlunch achgroup
     stanmath month0 month7 probsolv matheff;
```


Summary Of Multiple Imputation Estimates

Analysis results from 20 imputed data sets

Parameter	Est.	*SE*	*z*	*p*
Intercept	50.20	0.17	287.86	< .001
MONTH slope	0.43	0.04	10.79	< .001
CONDITION slope	-0.29	0.22	-1.30	0.194
MONTH x CONDITION	0.35	0.05	6.78	< .001

Interpretations

The mean problem-solving score at the beginning of the year for students in the control condition is $\hat{\beta}_0 = 50.20$

The initial status mean difference for students in the intervention condition is $\hat{\beta}_2 = -.29$ lower ($p = .19$)

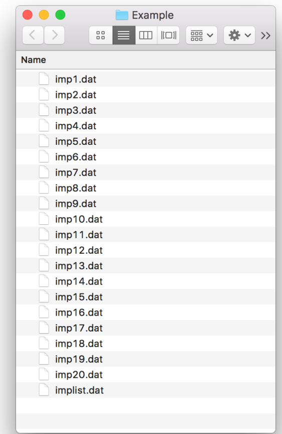
The average monthly growth rate for students in the control condition is $\hat{\beta}_1 = .43$ ($p < .001$)

The monthly growth rate for students in the intervention condition is $\hat{\beta}_3 = .35$ higher, on average ($p < .001$)

Mplus Imputation Format

Mplus requires imputed data sets as separate files

Blimp creates a text file containing the names of the data sets, and this file serves as the input data for subsequent analyses



Mplus Imputation Analysis Script

```
DATA:
file = implist.csv;
type = imputation;
VARIABLE:
names = school student wave condition eslpct ethnic male
  frlunch achvgrp stanmath month0 month7 probsolv matheff;
usevariables = month0 male frlunch stanmath condition probsolv;
cluster = student;
within = month0;
between = male frlunch stanmath condition;
```


Mplus Imputation Analysis Script

```
DEFINE:
center male frlunch stanmath (grandmean);
ANALYSIS:
type = twolevel random;
MODEL:
%within%
slope_i | probsolv on month0;
%between%
slope_i on condition;
probsolv on male frlunch stanmath condition;
probsolv with slope_i;
```

Mplus Output

```
MODEL FIT INFORMATION

Number of Free Parameters                        11

...

Wald Test of Parameter Constraints

          Value                               1484.785
          Degrees of Freedom                         6
          P-Value                               0.0000
```

61

Mplus Output

```
MODEL RESULTS
                                                    Two-Tailed   Rate of
                    Estimate       S.E.  Est./S.E.    P-Value    Missing
Within Level

 Residual Variances
    PROBSOLV          12.543      0.361     34.785      0.000      0.132

Between Level

 SLOPE_I    ON
    CONDITION          0.351      0.052      6.780      0.000      0.140

 PROBSOLV   ON
    MALE              -0.002      0.194     -0.010      0.992      0.087
    FRLUNCH           -0.156      0.245     -0.638      0.523      0.199
    STANMATH           0.026      0.001     26.202      0.000      0.048
    CONDITION         -0.290      0.223     -1.300      0.194      0.073

 PROBSOLV WITH
    SLOPE_I            0.017      0.104      0.167      0.867      0.172

 Intercepts
    PROBSOLV          50.198      0.174    287.864      0.000      0.093
    SLOPE_I            0.433      0.040     10.793      0.000      0.188

 Residual Variances
    PROBSOLV           5.237      0.557      9.408      0.000      0.173
    SLOPE_I            0.115      0.031      3.702      0.000      0.206
```

62

R Imputation Analysis Script

```
library(mitml)
library(plyr)
library(lme4)

# read stacked imputation data
setwd(dirname(rstudioapi::getActiveDocumentContext()$path))
impdata <- read.table(paste0(getwd(),"/imps_stacked.dat"))
names(impdata) <- c("imputation", "school", "student", "wave", "condition", "eslpct", “ethnic",
  “male", "frlunch", "achvgrp", “stanmath”, "month0", "month7", "probsolv", "matheff")
impdata <- impdata[impdata$imputation > 0, ]

# center covariates at grand means
impdata <- ddply(impdata, c("imputation"), transform, malec = scale(male, center = T, scale = F))
impdata <- ddply(impdata, c("imputation"), transform, frlunchc = scale(frlunch, center = T, scale = F))
impdata <- ddply(impdata, c("imputation"), transform, stanmathc = scale(stanmath, center = T, scale = F))
```

63

R Imputation Analysis Script

```
# analysis and pooling
implist <- as.mitml.list(split(impdata, impdata$imputation))
model <- "probsolv ~ month0  + condition + month0*condition + malec +
  frlunchc + stanmathc +  (month0 | student)"
analysis <- with(implist, lmer(model, REML = F))
estimates <- testEstimates(analysis, var.comp = T, df.com = NULL)
estimates

# estimate empty model with no predictors
emptymodel <- with(implist, lmer(probsolv ~ 1 + (month0 | student)))

# compare models with Wald test (e.g., MI version of omnibus F test)
testModels(analysis, emptymodel, method = "D1")

# compare models with likelihood ratio test (e.g., MI version of chi-square diff)
testModels(analysis, emptymodel, method = "D3")
```

64

R Output

```
Final parameter estimates and inferences obtained from 20 imputed data sets.

                 Estimate Std.Error  t.value       df  P(>|t|)      RIV      FMI
(Intercept)        50.198     0.179  279.954 2278.013    0.000    0.101    0.092
month0              0.433     0.042   10.257  675.301    0.000    0.202    0.170
condition          -0.290     0.227   -1.280 3798.174    0.201    0.076    0.071
malec              -0.002     0.191   -0.010 2430.850    0.992    0.097    0.089
frlunchc           -0.156     0.253   -0.617  567.744    0.537    0.224    0.186
stanmathc           0.026     0.001   28.048 6426.785    0.000    0.057    0.055
month0:condition    0.351     0.053    6.652 1071.917    0.000    0.154    0.135

                               Estimate
Intercept~~Intercept|student      5.236
Intercept~~month0|student         0.018
month0~~month0|student            0.115
Residual~~Residual               12.543
ICC|student                       0.294

Unadjusted hypothesis test as appropriate in larger samples.
```


R Output

```
Model comparison calculated from 20 imputed data sets.
Combination method: D1

    F.value      df1      df2    P(>F)      RIV
    243.560        6 8552.926    0.000    0.126

Unadjusted hypothesis test as appropriate in larger samples.


Model comparison calculated from 20 imputed data sets.
Combination method: D3

    F.value      df1      df2    P(>F)      RIV
    170.722        6 5397.762    0.000    0.164

Models originally fit with REML were automatically refit using ML.
```


SAS Imputation Analysis Script

```
/* read stacked imputation data */
data impdata (where = (_imputation_ gt 0));
infile '/folders/myfolders/imps_stacked.dat';
input _imputation_ school student wave condition eslpct ethnic male frlunch achgroup stanmath month0 month7
  probsolv matheff; run;

/* center covariate at grand means in each data set */
proc means data = impdata noprint;
var male frlunch stanmath;
by _imputation_;
output out = grandmeans (drop =_type_ _freq_) mean = malemean frlunchmean stanmathmean; run;

data impdata;
merge impdata grandmeans;
by _imputation_;
malec = male - malemean;
frlunchc = frlunch - frlunchmean;
stanmathc = stanmath - stanmathmean;  run;
```


SAS Imputation Analysis Script

```
/* analyze imputations */
ods _all_ close;
proc mixed data = impdata noclprint;
model probsolv = month0 condition month0*condition malec frlunchc stanmathc / solution covb;
random intercept month0 / subject = student type = un;
by _imputation_;
ods output SolutionF = estimates CovB = covb;
run;
ods listing;

/* pool estimates and standard errors */
proc mianalyze parms = estimates covb(effectvar = rowcol) = covb;
modeleffects Intercept month0 condition month0*condition malec frlunchc stanmathc;
run;
```

SAS Output

The MIANALYZE Procedure

Model Information	
PARMS Data Set	WORK.ESTIMATES
COVB Data Set	WORK.COVB
Number of Imputations	20

Variance Information (20 Imputations)

	Variance				Relative	Fraction	
Parameter	Between	Within	Total	DF	Relative Increase in Variance	Fraction Missing Information	Relative Efficiency
Intercept	0.002684	0.029338	0.032156	2473.7	0.096058	0.088377	0.995601
month0	0.000285	0.001485	0.001784	677.64	0.201124	0.169893	0.991577
condition	0.003462	0.047968	0.051604	3828.4	0.075787	0.070933	0.996466
month0*condition	0.000354	0.002423	0.002794	1075.8	0.153265	0.134504	0.993320
malec	0.003087	0.033586	0.036827	2453.4	0.096494	0.088745	0.995582
frlunchc	0.011161	0.052612	0.064332	572.51	0.222754	0.185016	0.990834
stanmathc	4.4314856E-8	0.000000813	0.000000860	6487.4	0.057214	0.054409	0.997287

Parameter Estimates (20 Imputations)

Parameter	Estimate	Std Error	95% Confidence Limits		DF	Minimum	Maximum	Theta0	t for H0: Parameter=Theta0	Pr > |t|
Intercept	50.197520	0.179320	49.84589	50.54915	2473.7	50.084016	50.280765	0	279.93	<.0001
month0	0.432869	0.042238	0.34994	0.51580	677.64	0.407317	0.469095	0	10.25	<.0001
condition	-0.290254	0.227164	-0.73563	0.15512	3828.4	-0.379308	-0.151923	0	-1.28	0.2014
month0*condition	0.351334	0.052861	0.24761	0.45506	1075.8	0.311484	0.380369	0	6.65	<.0001
malec	-0.001991	0.191904	-0.37830	0.37432	2453.4	-0.085553	0.114324	0	-0.01	0.9917
frlunchc	-0.156213	0.253637	-0.65439	0.34196	572.51	-0.347178	0.030670	0	-0.62	0.5382
stanmathc	0.025945	0.000927	0.02413	0.02776	6487.4	0.025520	0.026249	0	27.98	<.0001

SPSS Imputation Analysis Script

```
* set working directory.
CD "YOUR-FILE-PATH".

* read stacked imputation data.
DATA LIST free file = "imps_stacked.dat"
 /imputation_ school student wave condition eslpct ethnic male frlunch achvgrp stanmath
   month0 month7 probsolv matheff.
MISSING VALUES all (999).

* center covariates at their grand means.
AGGREGATE
 /break = imputation_
 /malemean = mean(male)
 /frlunchmean = mean(frlunch)
 /stanmathmean = mean(stanmath).
```


SPSS Imputation Analysis Script

```
COMPUTE malec = male - malemean.
COMPUTE frlunchc = frlunch - frlunchmean.
COMPUTE stanmathc = stanmath - stanmathmean.

* initiate pooling routines.
SORT CASES by imputation_.
SPLIT FILE layered by imputation_.

* analysis and pooling.
MIXED probsolv with month0 condition malec frlunchc stanmathc
 /print = testcov solution
 /fixed = intercept month0 condition month0*condition malec frlunchc stanmathc
 /random = intercept month0 | subject(student) covtype(un).
```


SPSS Analysis Output

Estimates of Fixed Effects[a]

							95% Confidence Interval	
imputation_	Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
.00	Intercept	50.263398	.190977	931.467	263.190	.000	49.888602	50.638194
	month0	.442461	.044698	900.548	9.899	.000	.354737	.530186
	condition	-.285675	.240175	887.468	-1.189	.235	-.757052	.185703
	month0 * condition	.330551	.055930	855.618	5.910	.000	.220774	.440327
	malec	.039058	.197563	857.359	.198	.843	-.348706	.426821
	frlunchc	-.147642	.246913	856.150	-.598	.550	-.632267	.336983
	stanmathc	.026317	.000982	853.837	26.798	.000	.024390	.028245
20.00	Intercept	50.288993	.171554	977.728	293.138	.000	49.952336	50.625650
	month0	.420813	.038486	980.000	10.934	.000	.345289	.496337
	condition	-.379308	.219346	978.915	-1.729	.084	-.809750	.051134
	month0 * condition	.368070	.049154	980.000	7.488	.000	.271611	.464529
	malec	.025561	.183489	977.000	.139	.889	-.334516	.385638
	frlunchc	.030670	.230628	977.000	.133	.894	-.421913	.483254
	stanmathc	.026103	.000908	977.000	28.747	.000	.024321	.027885
Pooled	Intercept	50.197517	.179653		279.414	.000	49.845217	50.549816
	month0	.432869	.042238		10.248	.000	.349935	.515803
	condition	-.290254	.227165		-1.278	.201	-.735629	.155121
	month0 * condition	.351334	.052861		6.646	.000	.247611	.455056
	malec	-.001991	.191904		-.010	.992	-.378302	.374320
	frlunchc	-.156213	.253637		-.616	.538	-.654386	.341959
	stanmathc	.025945	.000927		27.981	.000	.024128	.027763

a. Dependent Variable: probsolv.

Stata Imputation Analysis Script

```
// set working directory
cd "YOUR-FILE-PATH"

// read stacked data
clear
infile imp school student wave condition eslpct ethnic male frlunch achvgrp
  stanmath month0 month7 probsolv matheff using “imps_stacked.dat”

// create unique row id within each data set
generate rownum = student*100 + wave

// recode missing data in original data (imp = 0)
recode condition-rownum (999 = .)
```


Stata Imputation Analysis Script

```
// center covariates
egen malemeans = mean(male), by(imp)
egen frlunchmeans = mean(frlunch), by(imp)
egen stanmathmeans = mean(stanmath), by(imp)
gen malec = male - malemeans
gen frlunchc = frlunch - frlunchmeans
gen stanmathc = stanmath - stanmathmeans

// convert to mi data , analyze and pool
mi import flong, m(imp) id(rownum) imputed(condition - stanmathc) clear
mi estimate, cmdok: mixed probsolv month0 condition c.month0#c.condition
  malec frlunchc stanmathc || student: month0, cov(uns) var
```


Stata Output

```
Multiple-imputation estimates                   Imputations       =         20
Mixed-effects ML regression                     Number of obs     =      6,874

Group variable: student                         Number of groups  =        982
                                                Obs per group:
                                                              min =          7
                                                              avg =        7.0
                                                              max =          7
                                                Average RVI       =     0.1484
                                                Largest FMI       =     0.2131
DF adjustment:   Large sample                   DF:     min       =     433.06
                                                        avg       =   1,769.75
                                                        max       =   6,383.50
Model F test:       Equal FMI                   F(   6, 8541.2)   =     243.52
                                                Prob > F          =     0.0000

------------------------------------------------------------------------------------
           probsolv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------------+---------------------------------------------------------------
             month0 |   .4328691   .0422026    10.26   0.000      .350005    .5157332
          condition |   -.289854   .2267414    -1.28   0.201    -.7344009     .154693
                    |
c.month0#c.condition |   .3513336   .0528144     6.65   0.000     .2477023    .4549648
                    |
              malec |  -.0019358   .1914702    -0.01   0.992    -.3773977    .3735261
           frlunchc |  -.1564329   .2531714    -0.62   0.537    -.6537025    .3408366
          stanmathc |   .0259479   .0009253    28.04   0.000      .024134    .0277617
              _cons |   50.22005   .1791557   280.32   0.000     49.86873    50.57137
------------------------------------------------------------------------------------
```