Paper AA-09-2015

Using SAS for the Longitudinal Analysis of Difference Scores

Brandy R. Sinco, MS, University of Michigan, Ann Arbor, MI
Edith Kieffer, PhD, University of Michigan, Ann Arbor, MI
Michael S. Spencer, PhD, University of Michigan, Ann Arbor, MI
Gloria Palmisano, BS, MA, CHASS Center, Detroit, MI
Gretchen Piatt, PhD, University of Michigan, Ann Arbor, MI
Michele Heisler, M.D., University of Michigan, Ann Arbor, MI

ABSTRACT

Background: When longitudinal data has no missing baseline values, analysis of difference scores is one method of normalizing the error terms, even if the original outcome variable is non-normal. Adjusting for the baseline value as a covariate enables estimation of difference scores, with adjustment for the starting value.

Objective and Methods: Derive the linear mixed model (LMM) for difference scores, which will include terms for time, treatment group, interaction between time and treatment, and baseline value.

Demonstrate how to use the SAS data step to prepare a dataset for longitudinal analysis of difference scores. Present a SAS macro that uses Proc Mixed for analysis of difference scores, with adjustment for the baseline values of treatment groups.

Derive the formulas for contrasts between change scores between treatment groups, adjusted for baseline. Show how to convert the contrast equations to SAS Estimate statements. Further, explain how the between-group contrasts can be adjusted for multiple comparisons.

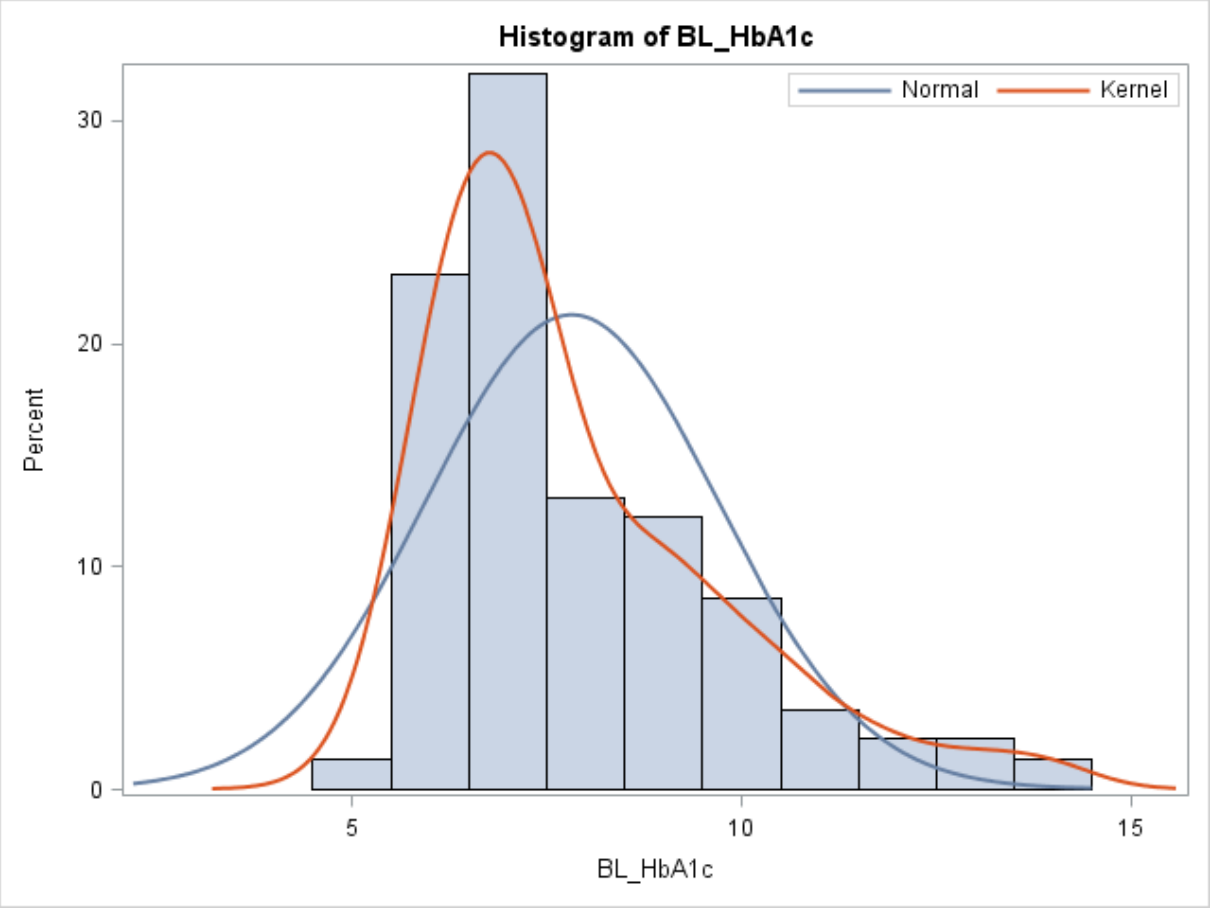
The example data will be from a diabetes study with three treatment groups with time points at baseline, 6-months, 12-months, and 18-months.

Results. Examples will be presented that show the trajectory of an outcome over time between treatment groups, in table and graphic format.
These will include the treatment group improving significantly, in comparison to the control group, and of the treatment group staying the same, while the control group worsened over time.

Conclusion. Outcome analysis, based on a LMM on difference scores with baseline adjustment, is an effective analysis technique for longitudinal data.

Outline

Figure 1: Histogram of Baseline Values With Kernal and Normal Densities



If the above variable needs to be analyzed as an outcome at baseline and follow-up time points, a transform is often necessary to reduce skewness (asymmetry) in the residuals. Common methods are log, square root, or other transforms. Today, let's focus on the difference score method to reduce skewness.

```
/* SAS Code for Figure 1 */
proc sgplot data=MyData;
  histogram BL_HbA1c;  density BL_HbA1c;  density BL_HbA1c / type=kernel;
  keylegend / location=inside position=topright; run;
```

Why Difference Scores Help to Reduce Skewness.

- Let Y_0= Outcome at baseline and Y_1 = Outcome at follow-up.
- Let Y_0 and Y_1 be random variables with means μ_0 and μ_1, variances σ_0^2 and σ_1^2, skewnesses γ_0 and γ_1, and Pearson correlation coefficient ρ.
-
- Then, the skewness of $(Y_1 - Y_0) = E((Y_1 - \mu_1) - (Y_0 - \mu_0))^3$
- $= E(Y_1 - \mu_1)^3 - 3E(Y_1 - \mu_1)^2(Y_0 - \mu_0)) + 3E((Y_1 - \mu_1)(Y_0 - \mu_0)^2) - E(Y_0 - \mu_0)^3$.
- If Y_0 and Y_1 are independent, the skewness of $(Y_1 - Y_0) = \gamma_1 - \gamma_0$, because the two expectation terms in the center will be zero.
- If Y_0 and Y_1 are not independent, the skewness of $(Y_1 - Y_0) = \gamma_1 - \gamma_0 + 3[E((Y_1 - \mu_1)(Y_0 - \mu_0)^2) - E(Y_1 - \mu_1)^2(Y_0 - \mu_0))]$.
- While there is no known formula for $E((Y_1 - \mu_1)(Y_0 - \mu_0)^2) - E(Y_1 - \mu_1)^2(Y_0 - \mu_0))]$, when the correlation coefficient is ρ, taking the difference score follow-up and baseline outcomes often helps to reduce skewness, because baseline and follow-up measures are often in the same family of distributions. Some examples are baseline and follow-up depression scores, blood pressures, and cholesterol measurements.

Baseline Adjustment in the Pre-Post Model With Proc GLM

- Let N = number of participants in a study, j = index of participant from 1 to N.
- Let G_k = group indicator, k = 0 control, 1 for treatment;
  G_k = 1 if person is a member of the group; 0 otherwise.
- Let Y_{j0} = outcome for the jth participant at baseline, aka pre-intervention or time 0.
- Let Y_{j1} = outcome for the jth participant at follow-up, or time 1.
- Let Δ_j = change score from baseline to follow-up = Y_{j1}- Y_{j0}.
- Let ε_j = error term.

Then, the change score model for a pre-post design would be
Let $\Delta j = \beta_0 + \beta_1 G_0 Y_{j0} + \beta_2 G_1 Y_{j0} + \beta_3 G_1 + \varepsilon_j$.

Let $\overline{Y}_{00}$ = baseline mean of control group.

Let $\overline{Y}_{10}$ = baseline mean of treatment group.

The estimated mean change score for the control group, adjusted for baseline = $\beta_0 + \beta_1 \overline{Y}_{00}$.

The estimated mean change score of the treatment group, adjusted for baseline = $\beta_0 + \beta_2 \overline{Y}_{10} + \beta_3$.

Intervention effect = Difference in change scores between treatment group, adjusted for baseline difference = $\beta_2 \overline{Y}_{10} - \beta_1 \overline{Y}_{00} + \beta_3$

When there is no missing data at baseline, the difference score model is an "intent-to-treat" model because it includes all available data at each time point.

The advantages to using the model with the difference score as the outcome, instead of Y_{j1} (value at time 1), are:
1)_The difference score often has a symmetrical error distribution, even when Y_{j1} (value at time 1) does not.

2)_To evaluate the effectiveness of a study, researchers are interested in the amount of change in the outcome. If the mean of the observed values at time 1 needs to be estimated, the baseline mean for the corresponding treatment group can easily be added.

3)_This model produces the average change within a treatment group, adjusted for the baseline value within that treatment group. Similarly, the intervention effect can be calculated to account for baseline difference between treatment groups.

For a real-life example, let's consider Hemoglobin A1c (HbA1c), the state of the art measure of blood sugar among people with diabetes.

SAS Code for Proc GLM to Test for Significant Change From Baseline to 6 Months in the Treatment and Control Groups

```
/* In a data step, compute difference score */
M6BL_HbA1c = M6_HbA1c – BL_HbA1c;

/* Then, compute Randomization, so that SAS will set default value 0 */
/* to the treatment group, because the largest value is the reference value */
RandomizationN = -Randomization;

/* Baseline mean HbA1c's = 7.9 treatment, 7.7 control */

ods html; ods graphics on;
Proc GLM Data= AcrossTimeHorizontal PLOTS =(RESIDUALS DIAGNOSTICS);
Class RandomizationN;  /* reference will be 0 = control */

Model M6BL_HbA1c=RandomizationN RandomizationN*BL_HbA1c  / Solution;

Estimate 'Treat M6 - BL'
       Intercept 1  RandomizationN 1 0 RandomizationN*BL_HbA1c 7.9 0;
```

```
Estimate 'Ctl M6 - BL'
      Intercept 1  RandomizationN 0 1 RandomizationN*BL_HbA1c 0 7.7;

Estimate 'Int Effect M6 - BL'
      RandomizationN 1 -1 RandomizationN*BL_HbA1c 7.9 -7.7;
Run; Quit;
ods graphics off; ods html close;
```

Table 1A: Proc GLM Output From Difference Score Model With Baseline Adjustment in Each Treatment Group

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Intercept	2.17	0.75	2.91	0.0041
RandomizationN -1	0.29	0.89	0.33	0.7427
RandomizationN 0	0	.	.	.
BL_HbA1c*Randomizati -1	-0.38	0.06	-6.09	<.0001
BL_HbA1c*Randomizati 0	-0.28	0.10	-2.94	0.0037

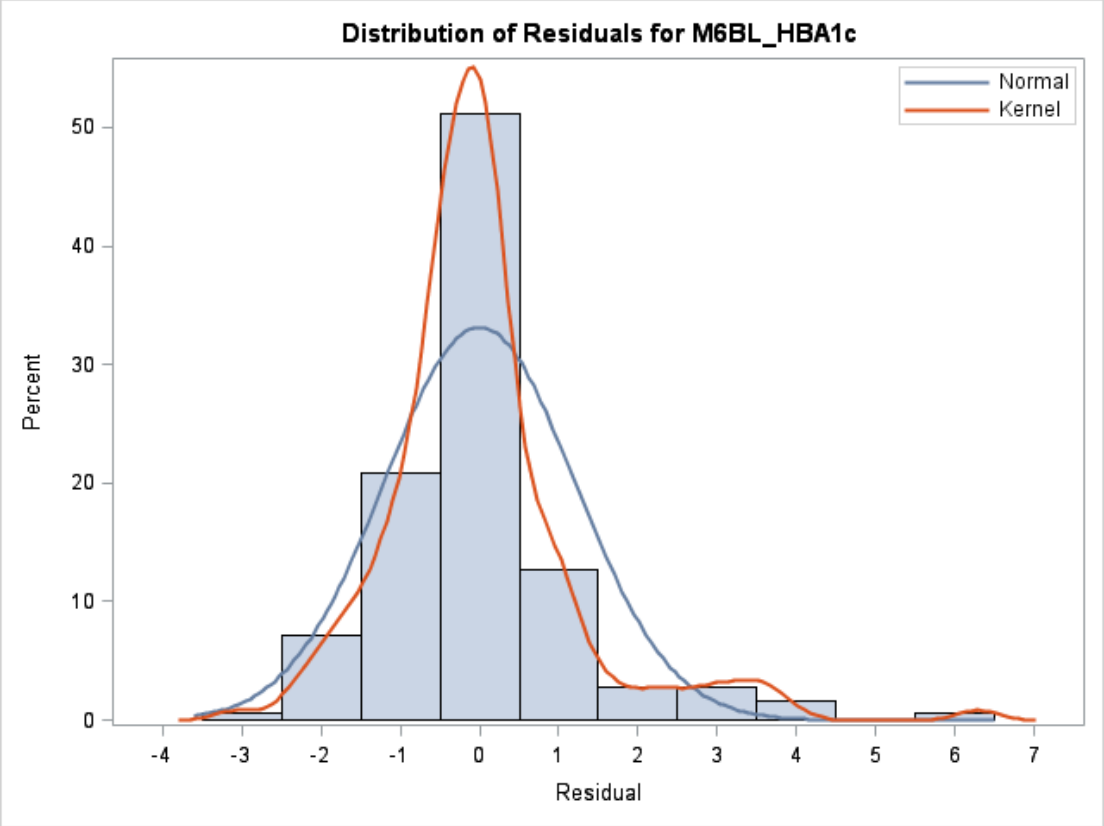
Note that RandomizationN, by itself, is non-significant. That is not a problem because the outcome of interest is the change score with baseline adjustment. Also, while the estimate for RandomizationN along is 0 for the control group, SAS estimates interaction effects between the baseline value and RandomizationN for both the treatment and control groups.

Table 1B: Proc GLM Estimates Difference Scores

Parameter	Estimate	Standard Error	t Value	Pr > |t|
Treatment M6 - BL	-0.53	0.11	-4.76	<.0001
Control M6 - BL	0.00	0.16	-0.01	0.9959
Intervention Effect M6 - BL	-0.53	0.19	-2.76	0.0064

In a diabetes intervention, a drop of at least 0.5 in HbA1c is considered clinically significant.
The above table indicates that the treatment group experienced a significant drop in HbA1c from baseline to 6-month follow-up, while the control group stayed approximately the same. The effect for the treatment group is significant within-group (1st line) and in comparison to the control group (3rd line).

Figure 2: Diagnostic Residual Plot for Month 6 – Baseline Difference Score



The above residuals are close to symmetric, although a little right-skewed.

SAS Macro to Test for Significant Change From Baseline to 6 Months in the Treatment and Control Groups

```
%Macro DiffM6(VarName, BLName, BLTreat, BLControl);
ods html; ods graphics on;
Proc GLM Data= AcrossTimeHorizontal PLOTS(only)=(RESIDUALS DIAGNOSTICS);
Class RandomizationN;
Model &VarName=RandomizationN RandomizationN*&BLName  / Solution;


Estimate 'Treat M6 - BL'  Intercept 1  RandomizationN 1 0 RandomizationN*&BLName
&BLTreat 0;
Estimate 'Control M6 - BL'  Intercept 1  RandomizationN 0 1 RandomizationN*&BLName 0
&BLControl;
Estimate 'Int Effect M6 - BL' RandomizationN 1 -1 RandomizationN*&BLName &BLTreat -
&BLControl;
Run; Quit;
ods graphics off; ods html close;
%MEnd DiffM6;

%DiffM6(M6BL_HbA1c, BL_HbA1c, 7.9, 7.7);
%DiffM6(M6BL_TotalCholesterol, BL_TotalCholesterol, 182.3, 182.9);
%DiffM6(M6BL_LDLCholesterol, BL_LDLCholesterol, 96.1, 95.4);
```

Introduction to the Linear Mixed Model for Longitudinal Data with Proc Mixed With Two Time Points and Two Treatment Groups

The general form of the linear mixed model is $Y = X\beta + Zb + \varepsilon$, where
X = matrix of fixed effects and Z = matrix of random effects,
B = fixed effect estimates, b = random effects estimates, and ε = error terms.
$b \sim N(0, D)$, $\varepsilon \sim N(0, \sum)$; b and ε are independent.[1]

For baseline and 6-month outcome data,

- Let Y_{ijk} = outcome variable; (i, j, k) = (randomization, time point, subject).
- i = 0 for control and 1 for treatment.
- j = 1 for pre-intervention and 2 for post-intervention.
- k = k^{th} subject.
- Let R = 0 for control and 1 for treatment.
- Let T = 0 for pre-intervention and 1 for post-intervention.

- Linear Mixed Model (LMM) for $Y_{ijk} = \beta_0 + \beta_1 R + \beta_2 T + \beta_3 RT + \varepsilon_{ijk}$, where ε_{ijk} = error term and $\varepsilon_{ijk} \sim N(0, \sum)$.
- Estimated Mean: $E(Y_{ij}) = \beta_0 + \beta_1 R + \beta_2 T + \beta_3 RT$.
- The means for the control group are β_0 at pre-intervention and $(\beta_0 + \beta_2)$ at post-intervention.
- The means for the treatment group are $(\beta_0 + \beta_1)$ at pre-intervention and $(\beta_0 + \beta_1 + \beta_2 + \beta_3)$ at post-intervention.
- The change scores are β_2 for the control group and $(\beta_2 + \beta_3)$ for the treatment group.
- The intervention effect is the difference in change scores for the treatment and control groups = β_3.

SAS Proc Mixed Code for Linear Mixed Model for Baseline and 6 Month Follow-Up

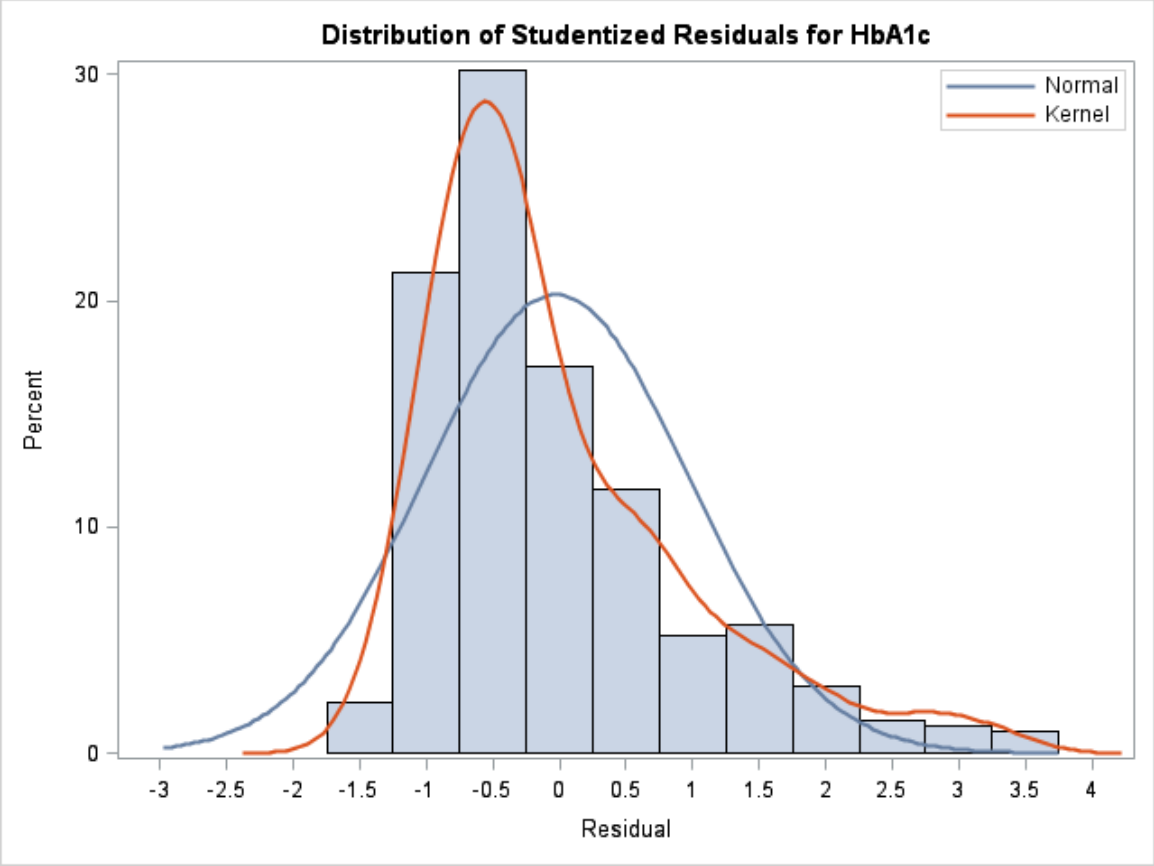
```
/* First, create vertical dataset from horizontal dataset */
/* Horizontal dataset has columns ID, BL_HbA1c (baseline value), */
/* M6_HbA1c (6-month value), M6BL_HbA1c (change score), */
/* Randomization (1=treatment, 0=control */
/* With one ID per record in the dataset. */
/* For longitudinal analysis, want the dataset in form: */
/* ID, Timepoint, Randomization, HbA1c (outcome value), delta_HbA1c (change score). */

Data Across;
Merge Baseline Month6Data Month12Data Month18Data;
by ID;

/* Compute difference scores, begin with HbA1c (blood sugar) */
M6BL_HbA1c=M6_HbA1c-BL_HbA1c;
M12BL_HbA1c=M12_HbA1c-BL_HbA1c;
M18BL_HbA1c=M18_HbA1c-BL_HbA1c;

M6BL_TotalCholesterol=M6_TotalCholesterol-BL_TotalCholesterol;
M12BL_TotalCholesterol=M12_TotalCholesterol-BL_TotalCholesterol;
M18BL_TotalCholesterol=M18_TotalCholesterol-BL_TotalCholesterol;

M6BL_LDLCholesterol=M6_LDLCholesterol-BL_LDLCholesterol;
M12BL_LDLCholesterol=M12_LDLCholesterol-BL_LDLCholesterol;
M18BL_LDLCholesterol=M18_LDLCholesterol-BL_LDLCholesterol;
Run;
```

```
/* Create Across_Long for Proc Mixed */
Data Across_Long;
Set Across;
/* Baseline */
Timepoint=0;
TimepointN=0;

HbA1c=BL_HbA1c;
TotalCholesterol=BL_TotalCholesterol;
LDLCholesterol=BL_LDLCholesterol;
Output; /* Need output statement at each time point */

/* 6 months */
Timepoint=1;
TimepointN=-1;

HbA1c=M6_HbA1c;
TotalCholesterol=M6_TotalCholesterol;
LDLCholesterol=M6_LDLCholesterol;

DeltaHbA1c=M6BL_HbA1c;
DeltaTotalCholesterol=M6BL_TotalCholesterol;
DeltaLDLCholesterol=M6BL_LDLCholesterol;
DeltaHDLCholesterol=M6BL_HDLCholesterol;
Output; /* Need output statement at each time point */

/* 12 Months */
Timepoint=2;
TimepointN=-2;

HbA1c=M12_HbA1c;
TotalCholesterol=M12_TotalCholesterol;
LDLCholesterol=M12_LDLCholesterol;

DeltaHbA1c=M12BL_HbA1c;
DeltaTotalCholesterol=M12BL_TotalCholesterol;
DeltaLDLCholesterol=M12BL_LDLCholesterol;
Output; /* Need output statement at each time point */
Run;

/* TimePointN = 0 for baseline and -1 for 6 months, coded so that SAS will set 0 as
the reference */
ods html; ods graphics on;
Proc Mixed Data= Across_Long Method=REML NOCLPRINT plots =(StudentPanel(conditional
box));

Class ID TimepointN RandomizationN;
Model HbA1c=TimepointN RandomizationN TimepointN*RandomizationN / Solution
Influence(effect=ID Est) ddfm=KR;

Repeated TimepointN / Type=UN Subject=ID R RCorr;

Estimate 'Treatment Baseline'
      Intercept 1 RandomizationN 1 0 TimePointN 0 1 TimepointN*RandomizationN 0 0 1
0;
Estimate 'Control Baseline'
      Intercept 1 RandomizationN 0 1 TimePointN 0 1 TimepointN*RandomizationN 0 0 0
1;
Estimate 'Treatment-Control Baseline'
      RandomizationN 1 -1 TimepointN*RandomizationN 0 0 1 -1;

Estimate 'Treatment Month 6'
```

```
      Intercept 1 RandomizationN 1 0 TimePointN 1 0 TimepointN*RandomizationN 1 0 0
0;
Estimate 'Control Month 6'
      Intercept 1 RandomizationN 0 1 TimePointN 1 0 TimepointN*RandomizationN 0 1 0
0;
Estimate 'Treatment M6 - BL'
      TimePointN 1 -1 TimepointN*RandomizationN 1 0 -1 0;
Estimate 'Control M6 - BL'
      TimePointN 1 -1 TimepointN*RandomizationN 0 1 0 -1;
Estimate 'Intervention Effect M6 - BL'
      TimepointN*RandomizationN 1 -1 -1 1;
Run;
ods graphics off; ods html close;
```

With the above model, the mean of any treatment group at any time point can be estimated.
However, the residual plot indicates departure from the normality assumption, as seen in Figure 2.

Figure 3: Diagnostic Residual Plot for Baseline, Month 6 Longitudinal Model



Longitudinal Difference Data at Four Time Points with Three treatment groups

Extension of the difference score model used with Proc GLM for baseline and follow-up.

- Extend to three treatment groups and four time points (baseline, 6 months, 12 months, 18 months).
- One application is to test whether the changes over time within treatment groups 1 or 2 continue to be significantly different from baseline after 6 months.

- Let N = number of participants in a study, j = index of participant from 1 to N.
- Let G_k = group indicator, k = 0 control, 1 for treatment 1, 2 for treatment.
  G_k = 1 if person is a member of the group; 0 otherwise.
- Let m = time point; 0 = baseline, 1 = 6 months, 2 = 12 months, 3 = 18 months.
- Let t_m = time point indicator; t_1 =1 if 6 months, 0 otherwise, t_2 = 1 if 12 and t_3 = 1 if 18 months.
- Let Y_{j0} = outcome for the jth participant at baseline, aka pre-intervention or time 0.
- Let Y_{jm} = outcome for the jth participant for follow-up time m.
- Let Δ_{jm} = change score from baseline to follow-up = Y_{jm}- Y_{j0}.
- Let ε_{jm} = error term.

The model equation will be

$$\Delta_{jm} = \beta_0 + \begin{bmatrix}\beta_1 & \beta_2\end{bmatrix}\begin{bmatrix}G_1 \\ G_2\end{bmatrix} + \begin{bmatrix}\beta_3 & \beta_4 & \beta_5\end{bmatrix}\begin{bmatrix}G_0Y_{j0} \\ G_1Y_{j0} \\ G_2Y_{j0}\end{bmatrix} + \begin{bmatrix}\beta_6 & \beta_7 & \beta_8\end{bmatrix}\begin{bmatrix}t_1 \\ t_2 \\ t_3\end{bmatrix} +$$

$$\begin{bmatrix}\beta_9 & \beta_{10} & \beta_{11} & \beta_{12} & \beta_{13} & \beta_{14}\end{bmatrix}\begin{bmatrix}G_1t_1 \\ G_1t_2 \\ G_1t_3 \\ G_2t_1 \\ G_2t_2 \\ G_2t_3\end{bmatrix} + \varepsilon_{jm}$$

- For example, consider the change in treatment group 1 from baseline to 12 months.
- Δ_group1_2 = $\beta_0 + \beta_1 + \beta_4\overline{Y}_{10} + \beta_7 + \beta_{10}$
- Change in group 2 from baseline to 12 months = $\beta_0 + \beta_2 + \beta_5\overline{Y}_{20} + \beta_7 + \beta_{13}$.
- Change in control group from baseline to 12 months = $\beta_0 + \beta_3\overline{Y}_{00} + \beta_7$.
- Intervention effect for group 1 at 12 months = $\beta_1 + \beta_4\overline{Y}_{10} - \beta_3\overline{Y}_{00} + \beta_{10}$.
- Average intervention effect for groups 1 and 2, compared to control =
  $.5\beta_1 + .5\beta_4\overline{Y}_{10} + .5\beta_{10} + .5\beta_2 + .5\beta_5\overline{Y}_{20} + .5\beta_{13} - \beta_3\overline{Y}_{00}$

- The SAS code for this model in Proc Mixed follows the same pattern as what we have seen before, except that an interaction between baseline value and treatment group must be added.

```
Proc Mixed Data= Across_Long Method=REML NOCLPRINT
plots(only)=(StudentPanel(conditional box));
Class ID TimepointN RandomizationN;
Model Delta_HbA1c=BL_HbA1c*RandomizationN TimepointN RandomizationN
TimepointN*RandomizationN / Solution Influence(effect=ID Est) ddfm=KR;
Repeated TimepointN / Type=UN Subject=ID;
Where TimePointN NE 0;
Run;
```

/* Macro Format, Estimate Statements Added */

```
*** Macro to calculate difference scores from baseline for 2 treatment groups,
compared to control group ***;
%Macro LMMDiff3(VarName, CorMat, BLName, BLRef, BLGrp1, BLGrp2);
/* VarName=outcome */
/* CorMat=Correlation structure */
/* BLName=Name of baseline variable */
/* BLRef=Baseline mean for the control or reference group */
/* BLGrp1=Baseline mean for treatment 1 group */
/* BLGrp2=Baseline mean for treatment 2 group */
ods html; ods graphics on;

*Compute average of treatment groups 1 & 2 to get combined intervention estimates;
%Let BLTrtAve=%sysfunc(mean(&BLGrp1, &BLGrp2));

*Use %syseval because SAS won't allow .5*macro variable in an estimate statement;
%Let BLTrtAve2=%sysevalf(&BLTrtAve/2);

Proc Mixed Data= Across_Long Method=REML NOCLPRINT
plots(only)=(StudentPanel(conditional box));;
Class ID TimepointN RandomizationN;
Model &VarName=&BLName*RandomizationN TimepointN RandomizationN
TimepointN*RandomizationN / Solution Influence(effect=ID Est) ddfm=KR;
Repeated TimepointN / Type=&CorMat Subject=ID R RCorr;


/* Estimate Examples at Month 12 */
Estimate 'Ref M12 - BL'  Intercept 1  RandomizationN 0 0 1 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 0 0 1 0 0 0 &BLName*RandomizationN 0 0 &BLRef;

Estimate 'Trt1 M12 - BL' Intercept 1  RandomizationN 0 1 0 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 0 1 0 0 0 0 &BLName*RandomizationN 0 &BLGrp1 0;

Estimate 'Trt2 M12 - BL' Intercept 1  RandomizationN 1 0 0 TimePointN 0 1 0
TimepointN*RandomizationN 0 0 0 1 0 0 0 0 0 &BLName*RandomizationN &BLGrp2 0 0;

Estimate 'Trt2-Trt1 M12-BL' RandomizationN 1 -1 0 TimepointN*RandomizationN 0 0 0 1 -1
0 0 0 0 &BLName*RandomizationN &BLGrp2 -&BLGrp1 0;

Estimate 'AveTrt1,2-RefM12-BL' RandomizationN .5 .5 -1 TimepointN*RandomizationN 0 0 0
.5 .5 -1 0 0 0 &BLName*RandomizationN &BLTrtAve2 &BlTrtAve2 -&BLRef;

Estimate 'Int Effect1 M12 - BL' RandomizationN 0 1 -1 TimepointN*RandomizationN 0 0 0
0 1 -1 0 0 0 &BLName*RandomizationN 0 &BLGrp1 -&BLRef;

Estimate 'Int Effect2 M12 - BL' RandomizationN 1 0 -1 TimepointN*RandomizationN 0 0 0
1 0 -1 0 0 0 &BLName*RandomizationN &BLGrp2 0 -&BLRef;

*** IMPORTANT: Exclude timepoint=0 rows ***;
Where TimePointN NE 0;
Run;
ods graphics off; ods html close;
%MEnd LMMDiff3;
```

Table 2A: Proc Mixed Output From Difference Score Model With Baseline Adjustment in Each Treatment Group

Effect	Time pointN	RandomizationN	Estimate	Std Err	Pr > |t|
Intercept			3.23	0.66	<.0001
BL_HbA1c*Randomizati		**-2**	-0.42	0.09	<.0001
BL_HbA1c*Randomizati		**-1**	-0.34	0.08	<.0001
BL_HbA1c*Randomizati		**0**	-0.43	0.08	<.0001
TimepointN	**-3**		0.28	0.24	0.2513
TimepointN	**-2**		0.03	0.16	0.8438
TimepointN	**-1**		0.00	.	.
RandomizationN		**-2**	-0.48	0.98	0.6252
RandomizationN		**-1**	-1.02	0.90	0.2621
RandomizationN		**0**	0.00	.	.
Timepoint*Randomizat	**-3**	**-2**	-0.11	0.34	0.7458
Timepoint*Randomizat	**-3**	**-1**	0.10	0.33	0.7638
Timepoint*Randomizat	**-3**	**0**	0.00	.	.
Timepoint*Randomizat	**-2**	**-2**	0.09	0.24	0.6895
Timepoint*Randomizat	**-2**	**-1**	0.21	0.22	0.326
Timepoint*Randomizat	**-2**	**0**	0.00	.	.
Timepoint*Randomizat	**-1**	**-2**	0.00	.	.
Timepoint*Randomizat	**-1**	**-1**	0.00	.	.
Timepoint*Randomizat	**-1**	**0**	0.00	.	.

Note: 0's in reference categories

However, our outcomes of interest are the difference scores, because there is no single beta coefficient that estimates change over time, adjusted for baseline.

Table 2B: Estimates of Difference Scores from Proc Mixed

Label	Estimate	Std Err	Pr > |t|
Reference M6 - BL	-0.06	0.16	0.691
Trt1 M6 - BL	**-0.36**	**0.15**	**0.014**
Trt2 M6 - BL	**-0.74**	**0.18**	**<.0001**
Trt2-Trt1 M6-BL	-0.38	0.23	0.097
AveTrt1,2- ReferenceM6-BL	**-0.48**	**0.20**	**0.016**
Int Effect1 M6 - BL	-0.30	0.21	0.166
Int Effect2 M6 - BL	-0.68	0.24	0.005
Reference M12 - BL	-0.03	0.17	0.853
Trt1 M12 - BL	-0.12	0.16	0.472
Trt2 M12 - BL	**-0.62**	**0.20**	**0.002**
Trt2-Trt1 M12-BL	-0.50	0.26	0.053
AveTrt1,2-ReferenceM12-BL	-0.32	0.22	0.135
Int Effect1 M12 - BL	-0.09	0.24	0.718
Int Effect2 M12 - BL	-0.59	0.26	0.027
Reference M18 - BL	0.21	0.24	0.380
Trt1 M18 - BL	0.02	0.23	0.946
Trt2 M18 - BL	**-0.58**	**0.24**	**0.020**
Trt2-Trt1 M18-BL	-0.59	0.34	0.082
AveTrt1,2-ReferenceM18-BL	-0.48	0.30	0.107
Int Effect1 M18 - BL	-0.20	0.34	0.559
Int Effect2 M18 - BL	**-0.79**	**0.34**	**0.023**

BL = baseline, M6 = month 6, M12 = month 12, M18 = month 18.
Trt = treatment group.
Int Effect = Intervention Effect.

Adjustment for Multiple Comparisons

In SAS Proc Logistic, Estimate statements can be written with the "Adjust=" option, which adjusts for multiple comparisons. Some of multiple comparison options are Tukey, Bonferroni, and Monte Carlo simulation. However, Procs GLM only allows the "Adjust=" option only on the LSMeans statement. Proc Mixed allows the Adjust=" option on both LSMeans and LSMEstimate statements, where all effects must be class variables. Because the baseline means are not categorical, adjustment for multiple comparisons can't be done with the estimate statements.

To adjust for multiple comparisons, use the model on page 7, in which the outcome is value at time point (i.e., HbA1c), rather than difference score (i.e., delta_HbA1c), and use LSMEstimate or LSMeans with the "Adjust=" option. The author uses the "Adjust=simulate" option, which adjusts by using Monte Carlo simulation, which performs well according to the literature[2].

Graphics and Tables from Publications

The longitudinal model described in this paper, with difference scores as the outcome and adjustment for baseline, has been used to publish two articles in medical journals. "Peer-led, empowerment-based, approach to self-management efforts in diabetes (PLEASED): A randomized controlled trial in the African-American community" was recently published in the Annals of Family Medicine[3]. "Comparative Effectiveness of Peer Leaders and Community Health Workers in Diabetes Self-management Support: Results of a Randomized Controlled Trial" was published Diabetes Care[4].

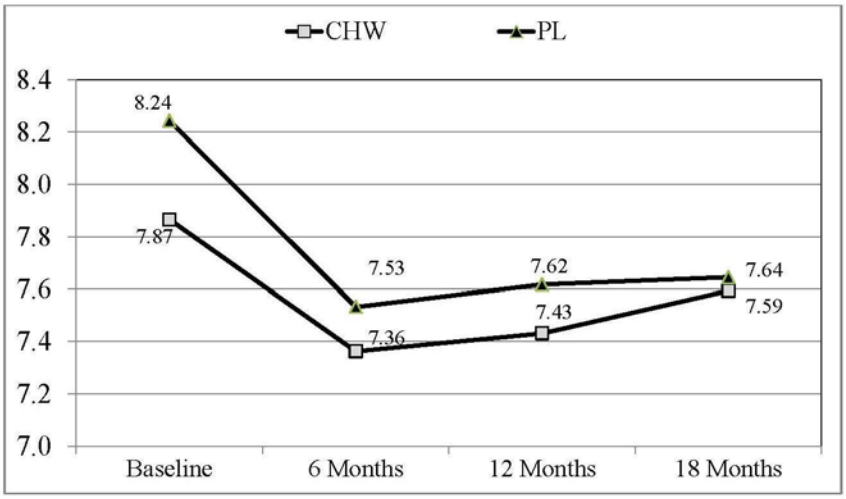
The table below was published in the Diabetes Care article[3]. This table compares blood sugar (HbA1c) drops across time between two groups. Participants in the CHW group received diabetes education and services from community health workers between baseline and 6-months. Whereas, the PL group received the same intervention as the CHW group, in addition to peer support between 6 months and 18 months. This analysis indicates that peer support may help to sustain the drop in HbA1c to 18 months. The drop in HbA1c continues to be significant in the PL group at 18 months at p<.05, although the means for the PL and CHW groups are not significantly different.

Table: Changes in Clinical and Psychological Outcomes Over Time Baseline – 18 Month Follow-Up (N=116: n=56 CHW, n=60 PL) Estimates for Means with 95% Confidence Intervals Linear Mixed Model for All Clinical Outcomes and Diabetes Support All Difference Scores Adjusted for Baseline Values

Outcome	Time Point	Baseline	6 Months – Baseline	12 Months – Baseline	18 Months – Baseline
HbA1c	**PL**	8.2 (7.7, 8.8)	-0.7 (-1.0, -0.4) p<.0001	-0.6 (-0.9, -0.3) p=0.001	-0.6 (-1.0, -0.2) p=0.009
	CHW	7.8 (7.4, 8.3)	-0.5 (-0.8, -0.3) p=0.0004	-0.4 (-0.7, -0.1) p=0.011	-0.3 (-0.7, 0.2) p=0.234
	CHW vs. PL**	0.253	0.883	0.867	0.725

The figure below is from the same article and shows the trajectory of HbA1c means over time

Figure 2: Trajectory of Unadjusted HbA1c Mean Scores Over Time



Conclusions.

- ✓ When there are no missing values in the outcome at baseline and the distribution is skewed, longitudinal analysis of differences scores can be a useful analysis technique.
- ✓ Taking the difference score between skewed variables from the same family of distributions often produces a result that is more symmetrical.
- ✓ Longitudinal analysis of difference scores can be implemented in SAS Proc GLM for two time points and in Proc Mixed for more than two time points.
- ✓ Analyses of longitudinal difference scores, with baseline adjustment, have been accepted for publication by major journals.
- ✓ Longitudinal analysis of difference scores is a useful technique in the data analyst's tool kit.

REFERENCES

1. West BT, Welch KB, Galecki AT. Linear mixed models: A practical guide using statistical software. 2nd ed. New York, NY: Chapman & Hall/CRC; 2014.

2. Westfall, P. H., and Young, S. S. (1993). Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment. New York: John Wiley & Sons.

3. Tang, T.S., Funnell, M.M., Sinco, B.R., Spencer, M.S., Heisler, M. Peer-led, empowerment-based, approach to self-management efforts in diabetes (PLEASED): A randomized controlled trial in the African-American community. Ann Fam Med August 2015 13:S27-S35; doi:10.1370/afm.1819.

4. Tang T.S., Funnell M., Sinco B., Piatt G., Palmisano G., Spencer M., Kieffer E., Heisler M. Comparative Effectiveness of Peer Leaders and Community Health Workers in Diabetes Self-management Support: Results of a Randomized Controlled Trial. Diabetes Care. 2014 Jun;37(6):1525-34. doi: 10.2337/dc13-2161.

CONTACT INFORMATION

Your comments and questions are welcome.

Brandy R. Sinco, Statistician and Programmer/Analyst
University of Michigan School of Social Work
1080 S. University St.
Box 183
Ann Arbor, MI 48109-1106

Phone: 734-763-7784

E-Mail: brsinco@umich.edu