PubH 7405:
REGRESSION ANALYSIS

Instructor: Chap T. Le, Ph.D.

Distinguished Professor of Biostatistics

INTRODUCTION TO REGRESSION &

THE ANALYSIS OF CORRELATED DATA

In all previous lectures, as well as in most standard statistical models, we assume that all observations collected were independent of each other. This last lecture introduces the exceptions with correlated data; you will see more of these next semester and beyond.

SPECIAL PROBLEMREGRESSION:
CORRELATED ERRORS
(Time Series Type)

AUTOCORRELATION

The basic multiple regression models have assume that the random error terms are independent normal random variables or, at least, uncorrelated random variables. In some fields – for example in economics, regression applications may involve "time series"; the assumption of uncorrelated or independent error terms may not be appropriate. In time series data, error terms are often (positively) correlated over time – auto-correlated or serially correlated.

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots \beta_k x_{ki} + \varepsilon_i$$

PROBLEMS OF AUTOCORRELATION

- Least squares estimates of regression coefficients are still unbiased but no longer have minimum variance
- MSE seriously under-estimate variance of error terms
- Standard errors of estimated regression coefficients may seriously under-estimate the true standard deviations of the estimated regression coefficients; confident intervals of regression coefficients and of response means, therefore, may not have the correct coverage.
- t and F tests may no longer applicable, have wrong size.

When the presence of autocorrelation is confirmed, the problem could be remedied by adding in another predictor variable or variables: one of the major cause is the omission from the model of one or more key predictors.

If nothing found, trying new models

AUTOREGRESSIVE ERROR MODEL

$$Y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots \beta_k x_{kt} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

$$where:$$

$$|\rho| < 1$$

$$u_i\text{'s are independen t } N(0, \sigma^2)$$

$$Y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots \beta_k x_{kt} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

$$where:$$

$$|\rho| < 1$$

$$u_i\text{'s are independen t } N(0, \sigma^2)$$

Note that each error term consists of a fraction of the previous error term plus a new disturbance term u_i; the parameter ρ – often positive - is called the "autocorrelation parameter".

First, we can easily expand from the definition of the first-order autoregressive error model to show that each error term is a linear combination of current and preceding disturbance terms. This is used to prove that the mean is zero and the variance is constant – BUT the variance is larger.

$$
\begin{aligned}
\varepsilon_t &= \rho\varepsilon_{t-1} + u_t \\
&= \rho(\rho\varepsilon_{t-2} + u_{t-1}) + u_t \\
&= \rho^2\varepsilon_{t-2} + \rho u_{t-1} + u_t \\
&= \rho^2(\rho\varepsilon_{t-3} + u_{t-2}) + \rho u_{t-1} + u_t \\
&= \rho^3\varepsilon_{t-3} + \rho^2 u_{t-2} + \rho u_{t-1} + u_t \\
&= \cdots \\
&= \sum_{s=0}^{\infty} \rho^s u_{t-s}
\end{aligned}
$$

$$
\begin{aligned}
E(\varepsilon_t) &= \sum_{s=0}^{\infty} \rho^s E(u_{t-s}) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
\sigma^2(\varepsilon_t) &= \sigma^2 \sum_{s=0}^{\infty} \rho^{2s} \\
&= \frac{\sigma^2}{1-\rho^2}
\end{aligned}
$$

The error terms of the first-order autoregressive model still have mean zero and constant variance but a positive covariance between consecutive terms:

$$Y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots \beta_k x_{kt} + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

where:

$$|\rho| < 1$$

u_i's are independen t $N(0, \sigma^2)$

$$E(\varepsilon_t) = 0$$

$$\sigma^2(\varepsilon_t) = \frac{\sigma^2}{1-\rho^2}$$

$$\boldsymbol{\sigma\{\varepsilon_t, \varepsilon_{t-1}\} = \rho\left(\frac{\sigma^2}{1-\rho^2}\right)}$$

The autocorrelation parameter is also the coefficient of correlation between two consecutive error terms:

$$\frac{\sigma(\varepsilon_t, \varepsilon_{t-1})}{\sigma(\varepsilon_t)\sigma(\varepsilon_{t-1})} = \frac{\rho\left(\frac{\sigma^2}{1-\rho^2}\right)}{\sqrt{\frac{\sigma^2}{1-\rho^2}}\sqrt{\frac{\sigma^2}{1-\rho^2}}} = \rho$$

The coefficient of correlation below, of two error terms that are s periods apart, shows that the error terms are also positively correlated but the further apart they are the less the correlation between them:

$$\frac{\sigma(\varepsilon_t, \varepsilon_{t-s})}{\sigma(\varepsilon_t)\sigma(\varepsilon_{t-s})} = \frac{\rho^s\left(\frac{\sigma^2}{1-\rho^2}\right)}{\sqrt{\frac{\sigma^2}{1-\rho^2}}\sqrt{\frac{\sigma^2}{1-\rho^2}}} = \rho^s$$

DURBIN-WATSON TEST

- **The Durbin-Watson test for autocorrelation assumes the first-order autoregressive error model (with values of predictor variables fixed)**
- **The test consists of determining whether or not the autocorrelation parameter (which is also the coefficient of correlation between consecutive error terms) is zero:**

$$H_0 : \rho = 0$$

$$H_A : \rho > 0$$

The Durbin-Watson test statistic D is obtained by first using ordinary least squares method to fit the regression function, calculating residuals and then D. Small values of D support the conclusion of autocorrelation because, under the first-order autoregressive error model, adjacent error terms tend to be of the same magnitude:

$$D = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

SAS implementation is very simple: Use option DW

PROC REG;

MODEL Y = X1 X2 X3/DW;

An estimate of the autocorrelation parameter is provided using the following formula:

$$r = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

FIXED EFFECTS, RANDOM EFFECTS, & MIXED EFFECTS

Consider a typical model for two-way ANOVA:

$$E(Y_{ij}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

$$\alpha = \text{Main Effect of A}$$

$$\beta = \text{Main Effect of B}$$

$$\alpha\beta = \text{Interaction Effect}$$

Suppose this is a randomized clinical trial where Factor A represents three treatments and Factor B represents genders (in a stratified randomization). This is the case of "Fixed Effects" where inferences are confined to the three specific treatments and the two genders.

Model assumptions are:

α's are constants subject to the restriction: $\Sigma\alpha = 0$
β's are constants subject to the restriction: $\Sigma\beta = 0$
αβ's are constants subject to the restriction: $\Sigma\alpha\beta = 0$

The F statistics are: MSA/MSE, MSB/MSE, and MSAB/MSE

$$E(MSA) = \sigma^2 + \frac{nb}{a-1}\sum \alpha_i^2$$

$$E(MSB) = \sigma^2 + \frac{na}{b-1}\sum \beta_i^2$$

$$E(MSAB) = \sigma^2 + \frac{n}{(a-1)(b-1)}\sum (\alpha\beta)_{ij}^2$$

$$E(MSE) = \sigma^2$$

Where a is the number of levels of Factor A, b is the number of levels of Factor B, and n the number of observations in each combination.

In a fixed effects model, all levels of a Factor are included in the study – with data. The effects are constants representing the differences between levels; when these constants/parameters are zero, the Factor has no effects. For example a study includes two doses (say 300mg and 600mg); Factor A (treatment) has two levels and we are interested only in the difference of these two specific doses; the winner would be selected for use.

Consider an investigation of the effects of "machine operator" (Factor A: 5 randomly selected operators; the factory has many operators) and "machine" (Factor B: 3 randomly selected machines; the factory has many machines). This is the case of "Random Effects" model where inferences are NOT confined to the 5 selected operators and 3 selected machines(but to "operators" and "machines" in general).

Model assumptions are:

α, β, αβ are independently and normally distributed as random variable with mean zero and variances σ_α^2, σ_β^2, and $\sigma_{\alpha\beta}^2$ respectively.

F statistics are MSA/MSAB, MSB/MSAB, and MSAB/MSE

$$E(MSA) = \sigma^2 + nb\sigma_{\alpha}^2 + n\sigma_{\alpha\beta}^2$$

$$E(MSB) = \sigma^2 + na\sigma_{\beta}^2 + n\sigma_{\alpha\beta}^2$$

$$E(MSAB) = \sigma^2 + n\sigma_{\alpha\beta}^2$$

$$E(MSE) = \sigma^2$$

Where a is the number of levels of Factor A, b is the number of levels of Factor B, and n the number of observations in each combination.

In a random effects model, only some levels of a Factor are included in the study – with data. The effects represent the variation among all levels – those included and those were not. When the "variance" representing the variation among levels is zero, the Factor has no effects. For example a study includes only 2 doses (say 300mg and 600mg); Factor A (treatment) has two levels represented but we are interested in the effects of "dose".

In some studies, one factor may have fixed effects, the other random effects; we have a "Mixed Effects" ANOVA model.

In regression, some terms/predictors may have fixed effects while others random effects; in these cases, we have a mixed effects regression model. The analysis is more complicated than that of ANOVA.

REPEATED MEASUREMENTS

In most standard statistical models, we assume that all observations collected were independent of each other. Two common types of studies violate that independence assumption:

(a) A longitudinal study, where participants are observed at several occasions over time, and

(b) A clustered study, where participants are inherently grouped in a way that responses measured from different members maybe correlated.

Longitudinal studies might be experimental, such as observing blood pressure over weeks after being randomly assigned to a new vs. standard anti-hypertensive medicine, or they might be observational, such as tracking fasting blood glucose over months in a population identified as being at high risk for diabetes. The *repeated measures are within person across time*.

Clustered studies might be experimental, such as a study that randomized an elementary school's classrooms to a new vs. the usual math curriculum and compared math test scores from students in those classrooms; each classroom is a cluster. Clustered studies might be observational, such as a study examining measures of dependence among elderly in several nursing homes; each nursing home is a cluster. The *repeated measures are within cluster across people*.

SIMPLE SOLUTION

For each of the studies with repeated measurements (longitudinal and clustered studies), a simple analysis approach would be to take the average of repeated measurements (or any other summary statistic) for each independent unit (a person or a cluster). Then applying standard statistical methods – such as t-test, ANOVA, or multiple regression.

In the longitudinal study of the new anti-hypertensive medication, we could calculate average blood pressure across the weeks of the study for each person, and then statistically compare the average (of the averages) in the new treatment group with the standard treatment group using a two-sample *t*-test or multiple regression to adjust for other subjects' characteristics.

In the longitudinal study of fasting blood glucose, we could calculate the slope in glucose (or area under the curve (AUC) for the glucose trajectory) for each person. We could then identify risk factors for having a positive slope (representing increasing fasting blood glucose) or a higher AUC (representing more exposure to glucose) using linear regression with the calculated slope or AUC as the response or Dependent Variable.

In the elementary school study, we could calculate the median test score per classroom and then statistically compare the averages (of medians) between new curriculum and usual curriculum classrooms with a non-parametric test.

These are all called *derived (or calculated) variable analyses* and are very simple to implement.

We used this derived/calculated variable analysis in the analysis of a two-period, two-treatment cross-over design in the investigation of Brussel Sprouts: (1) we average the 3 measurements in the same period, and (2) we subtract the averages (from the same subject) in two periods.

But this approach is not problem-free.

For example, in the glucose study, how many observations across time do we have for each person? Even in a well-designed study, the numbers of data points are not the same across people (missing data are quite common in longitudinal studies). In clustered studies, the cluster size varies from cluster to cluster.

As a result, in derived variable analyses, the variance of the (calculated) response is not constant; varies from subject to subject.

Among other reasons, these are not the type of heteroscedasticity that you could alleviate with weighted least squares because different variances are caused by different number of data points – not because the variance is a function of predictors (for which you can search for a "variance function" before setting weights).

STANDARD REGRESSION

In the standard multiple linear regression model, we have the mean of the response Y_i which is stipulated as a linear combination of the covariates x_{ji} taken on $i = 1,...,n$ subjects:

$$Y_i = \beta_0 + \sum_{k=1}^{K} \beta_k x_{ki} + \varepsilon_i$$

ε_i is a random error term (hence independent across the values of i) assumed to have a normal distribution with mean zero and constant variance σ^2. A regression coefficient β_k represents how much larger the mean response is with each 1-unit higher value for x_k.

We need one more layer of indexing to capture the multiple observations within subject; in a longitudinal study, we can think of a subject as a participant, while in a clustered study, we can think of a subject as a cluster:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + \varepsilon_{ij}^{*}$$

Here *i* continues to index subject (participant or cluster), *j* indexes observations taken within subject, and *k* indexes covariates.

$$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + \varepsilon_{ij}^{*}$$

For ε_{ij}^{*} , we still assume independence across the values of *i*, but we need to build in correlation across the values of *j* for each *i*; the asterisk is used in the notation only to make a distinction between the ε* in this formula and the ε in the next formula. A *linear mixed model* takes the approach of splitting ε* into components that represent variation between subjects and variation within subjects, similar to the motivation for a one-way ANOVA.

A SIMPLE LINEAR MIXED MODEL

This is the simplest linear mixed model:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + b_{i0} + \varepsilon_{ij}$$

This is called a *random intercept model*. We assume the b_{i0} are independent across *i* and normally distributed with mean zero and variance σ_0^2 (representing variation within independent units), while the ε_{ij} are independent across *i* and *j* and are normally distributed with mean zero and variance σ_e^2 (representing variation across independent units)

$$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + b_{i0} + \varepsilon_{ij}$$

The b_{i0} are thought of as "intercepts" because they are not multiplied by a covariate, but they are subject-specific: each subject has its own intercept b_{i0} that is added to what we call the population intercept, β_0. Since the b_{i0} are centered at zero due to the normality assumption, a subject with a large positive b_{i0} has larger response values on average than a subject with a small or negative b_{i0}.

$$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k x_{kij} + b_{i0} + \varepsilon_{ij}$$

How does using these additive, independently distributed normal components for the random intercept and the error induce correlation among the observations taken with a subject?

Consider a simple longitudinal example where our only covariate is a binary indicator for treatment group (x_1=1 if subject *i* is in the treatment group and x_1=0 if subject *i* is in the placebo group):

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + b_{i0} + \varepsilon_{ij}$$

$$Y_{ij} = \beta_0 + \beta_1 x_{1ij} + b_{i0} + \varepsilon_{ij}$$

Now suppose there are only two time points of measurement: *j*=1 for baseline and *j*=2 for follow-up. What is the covariance between those two measurements for a subject in the treatment group?

$$\begin{aligned} Cov(Y_{i1}, Y_{i2}) &= Cov(\beta_0 + \beta_1 + b_{i0} + \varepsilon_{i1}, \beta_0 + \beta_1 + b_{i0} + \varepsilon_{i2}) \\ &= Cov(b_{i0} + \varepsilon_{i1}, b_{i0} + \varepsilon_{i2}) \\ &= Cov(b_{i0}, b_{i0}) \end{aligned}$$

which by definition is equal to Var(b_{i0})= σ_0^2

$$Corr\left(Y_{i1}, Y_{i2}\right) = \frac{Cov\left(Y_{i1}, Y_{i2}\right)}{\sqrt{Var\left(Y_{i1}\right)Var\left(Y_{i2}\right)}} = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2}$$

This is sometimes called the *intra-class correlation* (ICC). The calculation would be the same for a subject in the placebo group. There is no *i* indexing either variance; the variances (and hence the ICC) are assumed to be the same for all subjects.

TESTING & INFERENCES

Estimation of the fixed effects ($\hat{\beta}_0$ and $\hat{\beta}_1$) and the variances, and prediction of the random effects ($\hat{b}_{i0}$), is done using a type of maximum likelihood estimation, specifically by maximizing the joint likelihood of two normally distributed quantities: the response conditional on the random effects ($Y_i|b_{i0}$) and the random effects (b_{i0}). Scientific interest usually lies in testing the fixed effects, and for this we use a *t* statistic where the degrees of freedom for the *t* statistic are determined from the sample size and the model that was fit

$$t = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

This above a “simple” model (one covariate). Suppose at baseline, researchers also recorded whether this was the woman’s first pregnancy and whether she had a history of anemia. The simple model could be expanded to form a “Multiple Regression Model” with two new fixed effects covariates.

If we are only interested in the primary outcome of “average serum ferritin, the “Random Intercept Model” would be adequate. It might be desirable to investigate a secondary outcome, the slope in serum ferritin over the six months for high dose compared to low dose; we would need to expand the model to include “Random Intercept” and “Random Slope”.

For example, the following is a linear mixed model with both random intercepts and random slope:

$$Y_{ij} = \beta_0 + \beta_1 TRT_{ij} + \beta_2 t_{ij} + \beta_3 (TRT_{ij})(t_{ij}) + \beta_4 PP_{ij} + \beta_5 ANEM_{ij} + b_{i0} + b_{i1} t_{ij} + \varepsilon_{ij}$$

In which *TR* is an indicator for high (TRT=1) versus low (TRT=0) dose; tij takes on the values 0, 1, 2, 3, 4, 5 for visit number; Var(εij) = ε_e^2, Var(b_{i0}) = σ_0^2, Var(b_{i1}) = σ_1^2 , and Cov(b_{i0},b_{i1}) = σ_{01}, and the random intercepts, slopes, and errors are all normally distributed and independent across I and *j* and of each other. We also include indicators for previous pregnancy (PP, yes/no) and history of anemia (ANEM, yes/no).

LME MODEL FOR CROSS-OVER DESIGN

There are two commonly used study designs in clinical research. In a **parallel study design**, each subject is randomly assigned to one and only one of two or several treatments. A **crossover design study** is a longitudinal study in which each subject receives a sequence of different treatments, and there is a washout period between two treatments.

Crossover designs are common for experiments in many scientific disciplines such as psychology, education, pharmaceutical science, and healthcare – especially medicine.

If the disease is chronic and the effect of the treatment is temporary/reversible, a crossover trial would be an attractive option.

SUMMARY

Design:

Group 1: Period #1 (A1) – washout – Period #2 (B2)

Group 2: Period #1 (B1) – washout – Period #2 (A2)

The letter is used to denote treatments (A or B) and the number, 1 or 2, denotes the period; e.g. "A1" for treatment in period #1. In the previous lecture/method, our data analysis is based on the following "outcome variables" (Treatment - Placebo):

X1 = A1 - B2; and X2 = A2 - B1

The subtraction of measurements from the same subject will mostly cancel or minimize effects of all subject-specific factors.

.

For crossover studies, the treatment effects from the earlier period might be carried over to the later period, which is called "carry-over effect". Due to this possible effect, the response might be changed over different period, so it could be referred to as an "interaction". We could not investigate it by the method in the previous lecture.

In this lecture, we explore a more powerful tool, the "linear mixed-effects" (LME) model. In this use, we model separately the treatment effects, the period effects, and the carryover effects or interaction

MODEL:

Define two indicator variables representing the Treatment (TRT) and Period (PRD) effects:

TRT = 1 for Treatment A and TRT = -1 for Treatment B

PRD = 1 for Period 1 and PRD = -1 for Period 2

And let X_{ijk} denote the response/observation for ith sequence, jth subject in kth period.

We consider this LME Model for Crossover Design:

$$X_{ijk} = \beta_0 + \beta_1 TRT + \beta_2 PRD + \beta_{12}(TRT)(PRD) + s_{ij} + \varepsilon_{ijk}$$

$$X_{ijk} = \beta_0 + \beta_1 TRT + \beta_2 PRD + \beta_{12}(TRT)(PRD) + s_{ij} + \varepsilon_{ijk}$$

The ‘s are fixed effects whereas sij is a term representing the random subject effects (a random intercept); s_{ij} is assumed to be independently and identically distributed as “normal” with mean zero, $N(0,\sigma_s^2)$. The last term, ε_{ijk}, denotes the error which is assumed to be independently and identically distributed as $N(0,\sigma^2)$.

$$X_{ijk} = \beta_0 + \beta_1 TRT + \beta_2 PRD + \beta_{12}(TRT)(PRD) + s_{ij} + \varepsilon_{ijk}$$

The inclusion of the random subject effects term, the random intercept, would capture the within-subject correlation, called the intra-class correlation, $\sigma_s^2/(\sigma_s^2+\sigma^2)$ – as seen previously.

From the definitions of TRT and PRD, as compared to the notation of the previous method/lecture (A1, A2, B1, and B2), we have:

$$E(A1) = \beta_0 + \beta_1 + \beta_2 + \beta_{12}$$

$$E(A2) = \beta_0 + \beta_1 - \beta_2 - \beta_{12}$$

$$E(A1) = \beta_0 - \beta_1 + \beta_2 - \beta_{12}$$

$$E(A1) = \beta_0 - \beta_1 - \beta_2 + \beta_{12}$$

$$E(A1) = \beta_0 + \beta_1 + \beta_2 + \beta_{12}$$

$$E(A2) = \beta_0 + \beta_1 - \beta_2 - \beta_{12}$$

$$E(A1) = \beta_0 - \beta_1 + \beta_2 - \beta_{12}$$

$$E(A1) = \beta_0 - \beta_1 - \beta_2 + \beta_{12}$$

$$E(X1 = A1 - B2) = 2(\beta_1 + \beta_2)$$

$$E(X2 = A2 - B1) = 2(\beta_1 - \beta_2)$$

$$\frac{E(X1) + E(X2)}{2} = 2\beta_1$$

$$\frac{E(X1) - E(X2)}{2} = 2\beta_2$$

If we compare these results to those in the previous method/lecture, the results on the Treatment Effects and Period Effects are the same in both methods.

The question is what is new with LME Model and, especially, what does the interaction term, β_{12}(TRT)(PRD), do?

TREATMENT EFFECTS

X1 is normally distributed as $N(\alpha+\beta,\sigma^2)$
X2 is normally distributed as $N(\alpha-\beta,\sigma^2)$

$\bar{x}_1$ is distributed as normal $N(\alpha+\beta, \frac{\sigma^2}{n})$,

$\bar{x}_2$ is distributed as normal $N(\alpha-\beta, \frac{\sigma^2}{n})$

$$a = \frac{\bar{x}_1 + \bar{x}_2}{2}$$

a is distributed as normal $N(\alpha(\frac{\sigma^2}{2n})$

ORDER EFFECTS

X1 is normally distributed as $N(\alpha+\beta,\sigma^2)$
X2 is normally distributed as $N(\alpha-\beta,\sigma^2)$

$\bar{x}_1$ is distributed as normal $N(\alpha+\beta, \frac{\sigma^2}{n})$,

$\bar{x}_2$ is distributed as normal $N(\alpha-\beta, \frac{\sigma^2}{n})$

$$b = \frac{\bar{x}_1 - \bar{x}_2}{2}$$

b is distributed as normal $N(\beta(\frac{\sigma^2}{2n})$

$$E(A1) = \beta_0 + \beta_1 + \beta_2 + \beta_{12}$$
$$E(A2) = \beta_0 + \beta_1 - \beta_2 - \beta_{12}$$
$$E(A1) = \beta_0 - \beta_1 + \beta_2 - \beta_{12}$$
$$E(A1) = \beta_0 - \beta_1 - \beta_2 + \beta_{12}$$

$$E(A1 - A2) = 2(\beta_2 + \beta_{12})$$

The difference between A1 (Treatment A in Period 1) and A2 (Treatment A in Period 2) is more than just Period effects; the interaction term,, represents the carried-over effects The use of LME Model allows us to investigate carried-over effects.

In crossover designs, one could make one single observation per period, or one could make multiple observations over different time points in the same period. If there are multiple observations per period, the choices are:

(1) Averaging these multiple observations, and us the previous model with a random intercept, or

(2) Including in the model and 3 new interaction terms (between Time and the previous 3 terms); these new four new terms have "random slopes.