PharmaSUG 2019 - Paper ST-321

PROC MIXED: Calculate Correlation Coefficients in the Presence of Repeated Measurements

Qinlei Huang, Merck & Co., Inc.
Radha Railkar, Merck & Co., Inc.

ABSTRACT

In drug and medical device development, it is often needed to evaluate the correlation between two technologies, platforms, assays, or devices, in a setting of a clinical trial with repeated measurements. The solution to this problem has been studied by many authors. Bland and Altman (1995) considered the problem by comparting the correlation into two components: between-subject and within-subject correlations. Lam, Webb and O'Donnell (1999) approached this problem by using maximum likelihood estimation in the case where the replicate measurements are linked over time. Roy (2006, 2015) solved the problem by considering different correlation structures on the repeated measures. This paper first reviews the statistical methods to estimate the Pearson's correlation coefficient between two measures in settings where multiple observations are available on the same subject; and then presents how to apply the CCRM (Correlation Coefficient for Repeated Measures) method using PROC MIXED in SAS to obtain the parameter estimates of interest. It includes the SAS example codes, as well as examples of hands-on data analysis and outputs.

Keywords: PROC MIXED, Correlation Coefficient, Repeated Measurements

INTRODUCTION

In clinical trials and pharmaceutical research for drug and medical device development, the correlation coefficient between two technologies and the comparison of measurements evaluated by different analytical techniques and assay platforms are often of interest. Pearson correlation coefficient assuming jointly normal distribution is the most widely used method to estimate correlation between two variables. Spearman correlation coefficient provides the non-parametric alternative based on ranks. However, neither method addresses the intraclass correlation within subject in the presence of repeated measurements in a longitudinal setting which may bias the correlation coefficient estimate. In this paper, we demonstrate how to use the CCRM (Correlation Coefficient for Repeated Measures) method, which is developed under the framework of correlation partitioning, to obtain an estimate of the correlation coefficient between two variables when repeated observations are available on each of the measures. A longitudinal setting is assumed when the two measures are linked over time.

CCRM: A MAXIMUM LIKELIHOOD-BASED METHOD UNDER THE FRAMEWORK OF CORRELATION PARTITIONING

How to evaluate the correlation between two variables in the presence of repeated measures has been studied by many scholars. Bland and Altman (1995a, b) demonstrated the analysis simply combining repeated observations can be highly misleading and tested different ad hoc solutions. They proposed that the true correlation coefficient ρ_{XY} between two variables with repeated measures can be partitioned into two components: the between-subject correlation ρ_{XYB} and the within-subject correlation ρ_{XYW}. The idea of correlation partitioning sets up the framework of a maximum likelihood-based method to calculate ρ_{XY}.

Lam (1999), Roy (2006, 2015), Hamlett (2003, 2004) and their colleagues developed and extended this method to different situations. Lam et al. (1999) derived the maximum likelihood estimates of ρ_{XYB} and ρ_{XYW} through a linear mixed-effects model assuming the repeated measures are linked over time. They started with the compound symmetry correlation structures between observations for each subject and later extended to a more complex auto-regressive correlation structure. Hamlett et al. (2003, 2004) explained how to analyze the data using PROC MIXED in SAS and generalized the model to settings where the repeated measurements are not linked over time. Roy et al. (2006, 2015) further illustrated the components of the between and within-subject correlation and generalized the model to various covariance patterns.

FRAMEWORK OF CORRELATION PARTITION: BETWEEN-SUBJECT AND WITHIN-SUBJECT CORRELATION

Under the framework of correlation partitioning, the true correlation coefficient can be considered to have two components: the between-subject correlation and the within-subject correlation. The between subject correlation measures the correlation between subject means and derives from the between-subject covariance matrix Σ_B. It indicates whether subjects with high average values of one variable (*x*) also tend to have high values of the other (*y*). The within-subject correlation assesses if an increase in one variable (*x*) within the individual is associated with an increase in the other (*y*). It derives from the within-subject covariance matrix Σ_W. The true correlation coefficient derives from the overall covariance matrix Σ, which is the combination of between-subject covariance matrix and within-subject covariance matrix.

Let (x_{ij}, y_{ij}) be the *j*th repeated observation (*j* = 1, …, m_i), a total of m_i repeated observations of the *x*, *y* variables taken from the *i*th subject (*i* = 1, …, n) in a sample of *n* individuals. Each observed pair (x_{ij}, y_{ij}) is assumed to be taken from a bivariate normal population with mean (μ_x, μ_y) and variance-covariance matrix Σ, i.e.

$$\begin{bmatrix} x_{ij} \\ y_{ij} \end{bmatrix} \sim N(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma)$$

where

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_Y\rho_{XY} \\ \sigma_X\sigma_Y\rho_{XY} & \sigma_Y^2 \end{bmatrix}$$

σ_X^2 and σ_Y^2 are the variances of *x* and *y* respectively, and ρ_{XY} is the true correlation between *x* and *y* to be estimated. For notational convenience, σ_{XY} will denote the covariance between *x* and *y* ($\sigma_{XY} = \sigma_X\sigma_Y\rho_{XY}$).

Following the definition of a linear mixed-effects model (Lam et al., 1999), the observed pair (x_{ij}, y_{ij}) can also be written as

$$x_{ij} = \mu_{xi} + e_{xij}$$

$$y_{ij} = \mu_{yi} + e_{yij}$$

where μ_{xi} and μ_{yi} are the true mean values of x and y for the *i*th individual, and e_{xij} and e_{yij} represent the within-subject variation. We assume that the subject effects μ_{xi} and μ_{yi} are random and independent of the random errors e_{xij} and e_{yij}, and they are bivariate normally distributed,

$$\begin{bmatrix} \mu_{xi} \\ \mu_{yi} \end{bmatrix} \sim N(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \Sigma_B)$$

where

$$\Sigma_B = \begin{bmatrix} \sigma_{XB}^2 & \sigma_{XB}\sigma_{YB}\rho_{XYB} \\ \sigma_{XB}\sigma_{YB}\rho_{XYB} & \sigma_{YB}^2 \end{bmatrix}$$

σ_{XB}^2 and σ_{YB}^2 are the between-subject variances of μ_{xi} and μ_{yi} correspondingly, and ρ_{XYB} is the true correlation between μ_{xi} and μ_{yi}. For notational convenience, σ_{XYB} will denote $\sigma_{XB}\sigma_{YB}\rho_{XYB}$.

Similarly, the random errors are assumed to have a bivariate normal distribution,

$$\begin{bmatrix} e_{xij} \\ e_{yij} \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma_W)$$

where

$$\Sigma_W = \begin{bmatrix} \sigma_{XW}^2 & \sigma_{XW}\sigma_{YW}\rho_{XYW} \\ \sigma_{XW}\sigma_{YW}\rho_{XYW} & \sigma_{YW}^2 \end{bmatrix}$$

σ_{XW}^2 and σ_{YW}^2 are the within-subject variances of e_{xij} and e_{yij} correspondingly, and ρ_{XYW} is the true within-subject correlation. For notational convenience, σ_{XYW} will denote $\sigma_{XW}\sigma_{YW}\rho_{XYW}$.

Thus, under the framework of correlation partitioning and following the definition of a linear mixed-effects model,

$$\Sigma = \Sigma_B + \Sigma_W = \begin{bmatrix} \sigma_{XB}^2 & \sigma_{XYB} \\ \sigma_{XYB} & \sigma_{YB}^2 \end{bmatrix} + \begin{bmatrix} \sigma_{XW}^2 & \sigma_{XYW} \\ \sigma_{XYW} & \sigma_{YW}^2 \end{bmatrix}$$

The true correlation coefficient between *x* and *y*, ρ_{XY}, can be derived based on a maximum likelihood function which simultaneously estimates both the between-subject and within-subject correlation.

$$\rho_{XY} = \frac{cov(x_{ij}, y_{ij})}{\sqrt{var(x_{ij})var(y_{ij})}} = \frac{\overbrace{cov[\mu_{xi}, \mu_{yi}]}^{between-subject\ covariance} + \overbrace{cov[(x_{ij} - \mu_{xi}), (y_{ij} - \mu_{yi})]}^{within-subject\ covariance}}{\sqrt{var(x_{ij})var(y_{ij})}} = \frac{\sigma_{XYW} + \sigma_{XYB}}{\sigma_X \sigma_Y}$$

CORRELATION STRUCTURE: TEMPORAL LINKS ACROSS TIME

Above section explains the specification and partitioning of covariance structure for the paired measures collected at the same time point. Full specification of the covariance structure requires assumptions regarding the relationship between X and Y measured at different time points. Here, the correlations between measurements taken at two different time points, j and j', j ≠ j' are defined as follows:

$$Corr(x_{ij}, x_{ij'}) = \rho_{XX'}$$

$$Corr(y_{ij}, y_{ij'}) = \rho_{YY'}$$

$$Corr(x_{ij}, y_{ij'}) = \rho_{XY}\delta$$

The correlations between measurements taken at two different time points j' and j'' are $\rho_{X'X''}$, $\rho_{Y'Y''}$, and $\rho_{XY}\delta'$. $\sigma_{XX'}$, $\sigma_{X'X''}$, $\sigma_{YY'}$, $\sigma_{Y'Y''}$, and $\sigma_{XY'}$, $\sigma_{X'Y''}$ will be used to denote corresponding covariance between measures taken at two different time points.

Accordingly, the assumed correlation structure is depicted in Figure 1.

$$\begin{array}{ccccccc}
\boldsymbol{x_{ij}} & \leftarrow & & \boldsymbol{\rho_{XY}} = \frac{\boldsymbol{\sigma_{XY}}}{\boldsymbol{\sigma_X^2\sigma_Y^2}} & & \rightarrow & \boldsymbol{y_{ij}} \\
\uparrow & \boxed{\nwarrow} & & & & \boxed{\nearrow} & \uparrow \\
\rho_{XjXj'} = \frac{\sigma_{XjXj'}}{\sigma_X^2} & & & \rho_{XY}\delta_{XjYj'} = \frac{\sigma_{XjYj'}}{\sigma_X^2\sigma_Y^2} & & & \rho_{YjYj'} = \frac{\sigma_{YjYj'}}{\sigma_Y^2} \\
\downarrow & \boxed{\swarrow} & & & & \boxed{\searrow} & \downarrow \\
\boldsymbol{x_{ij'}} & \leftarrow & & \boldsymbol{\rho_{XY}} = \frac{\boldsymbol{\sigma_{XY}}}{\boldsymbol{\sigma_X^2\sigma_Y^2}} & & \rightarrow & \boldsymbol{y_{ij'}} \\
\uparrow & \boxed{\nwarrow} & & & & \boxed{\nearrow} & \uparrow \\
\rho_{Xj'Xj''} = \frac{\sigma_{Xj'Xj''}}{\sigma_X^2} & & & \rho_{XY}\delta_{Xj'Yj''} = \frac{\sigma_{Xj'Yj''}}{\sigma_X^2\sigma_Y^2} & & & \rho_{Yj'Yj''} = \frac{\sigma_{Yj'Yj''}}{\sigma_Y^2} \\
\downarrow & \boxed{\swarrow} & & & & \boxed{\searrow} & \downarrow \\
\boldsymbol{x_{ij''}} & \leftarrow & & \boldsymbol{\rho_{XY}} = \frac{\boldsymbol{\sigma_{XY}}}{\boldsymbol{\sigma_X^2\sigma_Y^2}} & & \rightarrow & \boldsymbol{y_{ij''}}
\end{array}$$

Figure 1. Correlation Structure Specification With Links Over Time

The full covariance matrix for the entire set of 2*mi* repeated measurements for the *i*th subject is provided below to better visualize the covariance structure:

$$V_i = Cov\begin{pmatrix} X_{i1} \\ Y_{i1} \\ X_{i2} \\ Y_{i2} \\ \cdot \\ \cdot \\ X_{imi} \\ Y_{imi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma_X^2} & \boldsymbol{\sigma_{XY}} & \sigma_X^2\rho_{X1X2} & \sigma_{XY}\delta_{X1Y2} & \cdot\ \cdot & \sigma_X^2\rho_{X1Xmi} & \sigma_{XY}\delta_{X1Ymi} \\ \boldsymbol{\sigma_{XY}} & \boldsymbol{\sigma_Y^2} & \sigma_{XY}\delta_{X1Y2} & \sigma_Y^2\rho_{Y1Y2} & \cdot\ \cdot & \sigma_{XY}\delta_{X1Ymi} & \sigma_Y^2\rho_{Y1Ymi} \\ \sigma_X^2\rho_{X1X2} & \sigma_{XY}\delta_{X1Y2} & \boldsymbol{\sigma_X^2} & \boldsymbol{\sigma_{XY}} & \cdot\ \cdot & \sigma_X^2\rho_{X2Xmi} & \sigma_{XY}\delta_{X2Ymi} \\ \sigma_{XY}\delta_{X1Y2} & \sigma_Y^2\rho_{Y1Y2} & \boldsymbol{\sigma_{XY}} & \boldsymbol{\sigma_Y^2} & \cdot\ \cdot & \sigma_{XY}\delta_{X2Ymi} & \sigma_Y^2\rho_{Y2Ymi} \\ \cdot\ \cdot & & \cdot\ \cdot & & \cdot\ \cdot & \cdot\ \cdot & \\ \cdot\ \cdot & & \cdot\ \cdot & & \cdot\ \cdot & \cdot\ \cdot & \\ \sigma_X^2\rho_{X1Xmi} & \sigma_{XY}\delta_{X1Ymi} & \sigma_X^2\rho_{X2Xmi} & \sigma_{XY}\delta_{X2Ymi} & \cdot\ \cdot & \boldsymbol{\sigma_X^2} & \boldsymbol{\sigma_{XY}} \\ \sigma_{XY}\delta_{X1Ymi} & \sigma_Y^2\rho_{Y1Ymi} & \sigma_{XY}\delta_{X2Ymi} & \sigma_Y^2\rho_{Y2Ymi} & \cdot\ \cdot & \boldsymbol{\sigma_{XY}} & \boldsymbol{\sigma_Y^2} \end{pmatrix}$$

Please note that the covariance matrix has a similar structure for each subject, except for the dimension.

FULL CORRELATION STRUCTURE: PARTITION AND TEMPORAL LINKS

Under the framework of correlation partitioning with the inclusion of links over time, we further specify the full variance covariance structure *Vi* using the below general form, which includes two Kronecker products of four design matrices:

$$V_i = J_i \otimes \Sigma_B + K_i \otimes \Sigma_W$$

where Σ_B is a 2 x 2 matrix representing the between-subject variance covariance structure for each paired observation (x_{ij}, y_{ij}), and Σ_W is a 2 x 2 matrix depicting the within-subject variance covariance structure for each paired observation (x_{ij}, y_{ij}) at a given time point. It is assumed that Σ_B and Σ_W are not dependent on a particular time point and the same for all time points. J_i is an *mi* x *mi* matrix with all elements unity, and K_i is an *mi* x *mi* matrix which reflects the correlation pattern of the repeated measures on a given response variable. Identity matrix I_i is a special case of K_i. $\otimes$ denotes the Kronecker product.

A full specification of V_i is present below for better visualization. Please note that it is assumed K_i is the same for both response variable *x* and *y*, and thus $\rho_{Wjj'} = \rho_{XjXj'W} = \rho_{YjYj'W} = \delta_{XjYj'W} = \delta_{Wjj'}$.

$$V_i = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_{XB}^2 & \sigma_{XYB} \\ \sigma_{XYB} & \sigma_{YB}^2 \end{pmatrix}$$

$$+ \begin{pmatrix} 1 & \rho_{W12} & \cdots & \rho_{W1mi} \\ \rho_{W12} & 1 & \cdots & \rho_{W2mi} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{W1mi} & \rho_{W2mi} & \dots & 1 \end{pmatrix} \otimes \begin{pmatrix} \sigma_{XW}^2 & \sigma_{XYW} \\ \sigma_{XYW} & \sigma_{YW}^2 \end{pmatrix}$$

$$= Cov\begin{pmatrix} X_{i1} \\ Y_{i1} \\ X_{i2} \\ Y_{i2} \\ \cdot \\ \cdot \\ X_{imi} \\ Y_{imi} \end{pmatrix} = \begin{pmatrix} \sigma_{XB}^2 & \sigma_{XYB} & \sigma_{XB}^2 & \sigma_{XYB} & \cdots & \sigma_{XB}^2 & \sigma_{XYB} \\ \sigma_{XYB} & \sigma_{YB}^2 & \sigma_{XYB} & \sigma_{YB}^2 & \cdots & \sigma_{XYB} & \sigma_{YB}^2 \\ \sigma_{XB}^2 & \sigma_{XYB} & \sigma_X^2 & \sigma_{XY} & \cdots & \sigma_X^2 & \sigma_{XY} \\ \sigma_{XYB} & \sigma_{YB}^2 & \sigma_{XY} & \sigma_Y^2 & \cdots & \sigma_{XY} & \sigma_Y^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{XB}^2 & \sigma_{XYB} & \sigma_{XB}^2 & \sigma_{XYB} & \cdots & \sigma_{XB}^2 & \sigma_{XYB} \\ \sigma_{XYB} & \sigma_{YB}^2 & \sigma_{XYB} & \sigma_{YB}^2 & \cdots & \sigma_{XYB} & \sigma_{YB}^2 \end{pmatrix}$$

$$+ \begin{pmatrix} \boldsymbol{\sigma_X^2} & \boldsymbol{\sigma_{XY}} & \sigma_{XW}^2\rho_{W12} & \sigma_{XYW}\delta_{W12} & \cdots & \sigma_{XW}^2\rho_{W1mi} & \sigma_{XYW}\rho_{W1mi} \\ \boldsymbol{\sigma_{XY}} & \boldsymbol{\sigma_Y^2} & \sigma_{XYW}\delta_{W12} & \sigma_{YW}^2\rho_{W12} & \cdots & \sigma_{XY}\rho_{W1mi} & \sigma_Y^2\rho_{W1mi} \\ \sigma_{XW}^2\rho_{W12} & \sigma_{XYW}\rho_{W12} & \boldsymbol{\sigma_X^2} & \boldsymbol{\sigma_{XY}} & \cdots & \sigma_X^2\rho_{W2mi} & \sigma_{XY}\delta_{W2mi} \\ \sigma_{XYW}\rho_{W12} & \sigma_{YW}^2\rho_{W12} & \boldsymbol{\sigma_{XY}} & \boldsymbol{\sigma_Y^2} & \cdots & \sigma_{XY}\delta_{W2mi} & \sigma_Y^2\rho_{W2mi} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{XW}^2\rho_{W1mi} & \sigma_{XYW}\rho_{W1mi} & \sigma_X^2\rho_{W2mi} & \sigma_{XY}\rho_{W2mi} & \cdots & \sigma_{XB}^2 & \sigma_{XYB} \\ \sigma_{XY}\rho_{W1mi} & \sigma_Y^2\rho_{W1mi} & \sigma_{XY}\rho_{W2mi} & \sigma_Y^2\rho_{W2mi} & \cdots & \sigma_{XYB} & \sigma_{YB}^2 \end{pmatrix}$$

LINEAR MIXED-EFFECTS MODEL

Linear mixed-effects models are an extension of general linear models. It allows deviations from the assumption of independent and identically distributed normal random errors by the inclusion of additional random-effects parameters and more flexible covariance matrix of the random errors. It accommodates both correlated errors and errors with heterogeneous variances.

Let $Y_i = [x_{i1}, y_{i1,} x_{i2}, y_{i2,} \dots \dots, x_{im1}, y_{imi,}]$ be the 2*mi* dimensional vector of observations on subject *i*, for each *i*th subject, the linear mixed-effects model is considered:

$$\mathrm{y}_i = \mathrm{X}_i \beta + \mathrm{Z}_i \gamma_i + \varepsilon_i$$

Where y_i denotes the vector of observed outcomes; X_i is the known matrix of fixed effects; β is the unknown vector of fixed-effects parameters; Z_i is the known matrix of random effects; γ_i is the unknown vector of random-effects parameters; and ε_i is the unobserved vector of random errors. Both γ and ε are normally distributed with

$$E\begin{bmatrix}\gamma_i \\ \varepsilon_i\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix} \qquad var\begin{bmatrix}\gamma_i \\ \varepsilon_i\end{bmatrix} = \begin{bmatrix}G_i & 0 \\ 0 & R_i\end{bmatrix}$$

The variance of y_i is, therefore, $\mathrm{V}_i = \mathrm{Z}_i \mathrm{G}_i \mathrm{Z}_i' + \mathrm{R}_i$. It is determined by the design matrix of random effects, Z_i, the covariance matrix of random-effects parameters, G_i, and the covariance matrix of random errors, R_i.

The covariance structure for the G_i matrix models the error that represents the natural heterogeneity between subjects (i.e., between-subject sources of variability). And we have,

$$G_i = \Sigma_B$$

The covariance structure in the R_i matrix models the serial correlations (i.e., within-subject sources of variability), which is directly related to the spacing of measurements.

$$R_i = K_i \otimes \Sigma_W$$

γ_i is a 2 x 1 matrix and Z_i is a 2*mi* x 2 matrix.

$$\gamma_i = \begin{bmatrix}\gamma_1 \\ \gamma_2\end{bmatrix} \qquad Z_i = \begin{pmatrix}1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1\end{pmatrix}$$

Hence we have,

$$V_i = J_i \otimes \Sigma_B + K_i \otimes \Sigma_W = Z_i G_i Z_i' + R_i$$

VARIANCE COVARIANCE STRUCTURE SPECIFICATIONS: FIVE SCENARIOS

In this section, we are going to discuss the specifications of the variance covariance structures $\boldsymbol{G_i}$ and $\boldsymbol{R_i}$ for the linear mixed-effects models. All models will be fit using the SAS PROC MIXED procedure.

Between-Subject Covariance Patterns

As shown in the above section, the between-subject matrix $G_i = \Sigma_B$ is a 2 x 2 matrix. The specification of G_i is primarily based on the two assumptions: if the between-subject variances for *x* (σ_{XB}^2) and *y* (σ_{YB}^2) are equal or not; and if the subject-specific *X* and *Y* random effects are correlated or not. The most widely used variance pattern for G_i is UN. It allows for different between-subject variances for *x* and *y* ($\sigma_{XB}^2 \neq \sigma_{YB}^2$) and assumes the subject-specific *X* and *Y* random effects are correlated ($\sigma_{XYB} \neq 0$). If the equal variance assumption is met (σ_{XB}^2=σ_{YB}^2), CS structure will be chosen for parsimonious purpose.

The RANDOM statement in PROC MIXED is used to define random effects and specify the structure of G matrix. The SUBJECT = option defines the blocks of G and the TYPE = option specifies the covariance structure of each block G_i.

Within-Subject Covariance Patterns

The within-subject matrix R_i is a 2*mi* x 2*mi* matrix which has two parts: Σ_W and K_i.

Σ_W is a 2 x 2 matrix which represents the within-subject variance covariance structure for paired observations measured at the same time points. The specification of Σ_W is primarily based on the two assumptions: if the within-subject variances for *x* (σ_{XW}^2) and *y* (σ_{YW}^2) are equal or not; and if the within-subject correlation between *x* and *y* (σ_{XYW}) exists. The most widely used variance pattern for Σ_W is UN. It allows for different between-subject variances for *x* and *y* ($\sigma_{XW}^2 \neq \sigma_{YW}^2$). If the equal variance assumption is met (σ_{XW}^2=σ_{YW}^2), CS structure will be chosen for parsimonious purpose. Both assume non-zero within-subject correlation between *x* and *y* ($\sigma_{XYW} \neq 0$).

K_i represents the m_i x m_i dimensional correlation matrix of the repeated measurements over time on a given method. The specification of K_i is primarily based on the assumption of the correlations ($\rho_{Wjj'}$) among repeated measures at different time points.

If correlations among repeated measures from different time point is zero or negligible, K_i is an identity matrix I_i with diagonal equal to 1 and off diagonal equal to 0, i.e, $\rho_{Wjj'} = 0$.

$$K_i = \begin{pmatrix} 1 & & & \cdots & \\ 0 & 1 & & \cdots & \\ 0 & 0 & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

If the correlations among all the repeated measures within a subject are equal, K_i has a CS structure, i.e, $\rho_{Wjj'} = \delta$

$$K_i = \begin{pmatrix} 1 & & & \cdots & \\ \delta_W & 1 & & \cdots & \\ \delta_W & \delta_W & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ \delta_W & \delta_W & \delta_W & \cdots & 1 \end{pmatrix}$$

If the correlations among the repeated measures decay over time, K_i has an AR(1) structure which defines $\rho_{Wjj'} = \delta^{|j-j'|}$ for $j \neq j'$,

$$K_i = \begin{pmatrix} 1 & & & \cdots & \\ \delta_W & 1 & & \cdots & \\ \delta_W^2 & \delta & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_W^{mi-1} & \delta_W^{mi-2} & \delta_W^{mi-3} & \cdots & 1 \end{pmatrix}$$

If the correlations among the repeated measures are different for any two different time points, K_i has an UN structure.

$$K_i = \begin{pmatrix} 1 & & & \cdots & \\ \delta_{W12} & 1 & & \cdots & \\ \delta_{W13} & \delta_{W23} & 1 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_{W1mi} & \delta_{W2mi} & \delta_{W3mi} & \cdots & 1 \end{pmatrix}$$

Please note that it is assumed K_i is the same for both response variable *x* and *y*, and thus $\rho_{Wjj\prime} = \rho_{XjXj\prime W} = \rho_{YjYj'W} = \delta_{XjYj\prime W} = \delta_{Wjj\prime}$.

The REPEATED statement in PROC MIXED is used to specify the R matrix for random errors in the linear mixed-effects model. The SUBJECT = option defines the blocks of R and the TYPE = option specifies the covariance structure of each block R_i. R requests that blocks of the estimated R matrix, R_i, be displayed. RCORR produces the correlation matrix corresponding to blocks of the estimated R matrix.

Five Scenarios

To better illustrate and compare the impact of the different correlation structure specifications on the correlation coefficient estimates, the below five scenarios will be further discussed in the next section. For simplification purpose, the between-subject variance covariance structure G_i is always specified with an UN structure. Corresponding example SAS codes will be presented.

1. $R_i = I_i \otimes \Sigma_W$, I_i is an identity matrix and Σ_W has a CS structure
2. $R_i = I_i \otimes \Sigma_W$, I_i is an identity matrix and Σ_W has a UN structure
3. $R_i = K_i \otimes \Sigma_W$, K_i has a CS structure and Σ_W has a UN structure
4. $R_i = K_i \otimes \Sigma_W$, K_i has an AR(1) structure and Σ_W has a UN structure
5. $R_i = K_i \otimes \Sigma_W$, K_i has an UN structure and Σ_W has a UN structure

PROC MIXED: CALCULATE CORRELATION COEFFICIENT IN THE PRESENCE OF REPEATED MEASURES

EXAMPLE DATA

Following Lam, Roy, and Hamlett, we use the data taken from Bland and Altman (1994) which shows a temporally linked data set of repeated measurements of intramural pH and $PaCO_2$ obtained from gastric pH and blood gas analysis of eight subjects. Note that the number of repeated measurements varies by patient. Table 1 shows the dataset.

Subject #	Time	pH	$PaCO_2$	Subject #	Time	pH	$PaCO_2$	Subject #	Time	pH	$PaCO_2$
1	1	6.68	3.97	3	9	7.34	5.07	6	3	7.29	5.12
1	2	6.53	4.12	4	1	7.36	5.67	6	4	7.33	4.93
1	3	6.43	4.09	4	2	7.33	5.10	6	5	7.31	5.03
1	4	6.33	3.97	4	3	7.29	5.53	6	6	7.33	4.93
2	1	6.85	5.27	4	4	7.30	4.75	7	1	6.86	6.85
2	2	7.06	5.37	4	5	7.35	5.51	7	2	6.94	6.44
2	3	7.13	5.41	5	1	7.35	4.28	7	3	6.92	6.52
2	4	7.17	5.44	5	2	7.30	4.44	8	1	7.19	5.28
3	1	7.40	5.67	5	3	7.30	4.32	8	2	7.29	4.56
3	2	7.42	3.64	5	4	7.37	3.23	8	3	7.21	4.34
3	3	7.41	4.32	5	5	7.27	4.46	8	4	7.25	4.32
3	4	7.37	4.73	5	6	7.28	4.72	8	5	7.20	4.41
3	5	7.34	4.96	5	7	7.32	4.75	8	6	7.19	3.69
3	6	7.35	5.04	5	8	7.32	4.99	8	7	6.77	6.09
3	7	7.28	5.22	6	1	7.38	4.78	8	8	6.82	5.58
3	8	7.30	4.82	6	2	7.30	4.73				

Table 1. Repeated measures of intramural pH and $PaCO_2$ from 8 subjects

For analysis purpose, the above data is restructured, and a SAS dataset `example` is created. It has 4 variables: `subject`, `time`, `method`, and `y`.

- `subject` is the identification variable of each subject
- `time` represents time points
- `method` indicates observation type. method = 1 indicates pH and method = 2 indicates $PaCO_2$.
- `y` is a continuous response variable

Table 2 shows parts of the SAS dataset `example`.

Subject	Time	Method	Y
1	1	1	6.68
1	1	2	3.97
1	2	1	6.53
1	2	2	4.12
…	…	…	…
8	8	1	6.82
8	8	2	5.58

Table 2. SAS Dataset Example

GRAPHICAL REPRESENTATION

The below scatterplots (Figure 2 and Figure 3) are generated to visualize (1) the correlations between the subject means of pH and $PaCO_2$, and (2) the correlations between pH and $PaCO_2$ for each subject.

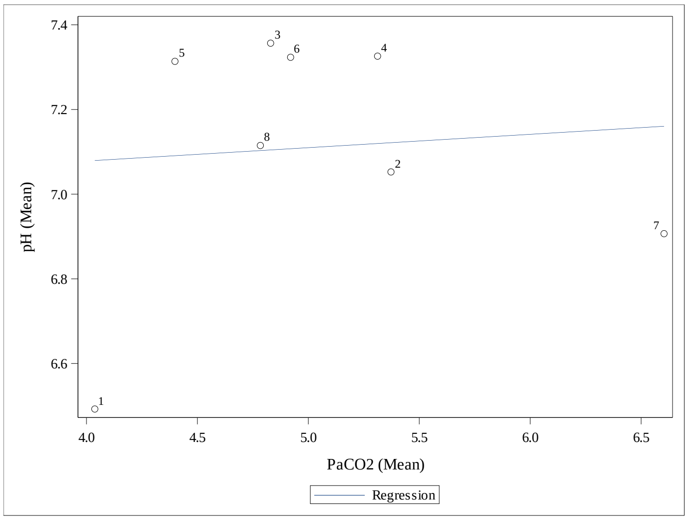

Figure 2. Between-Subject Correlations Between pH and $PaCO_2$

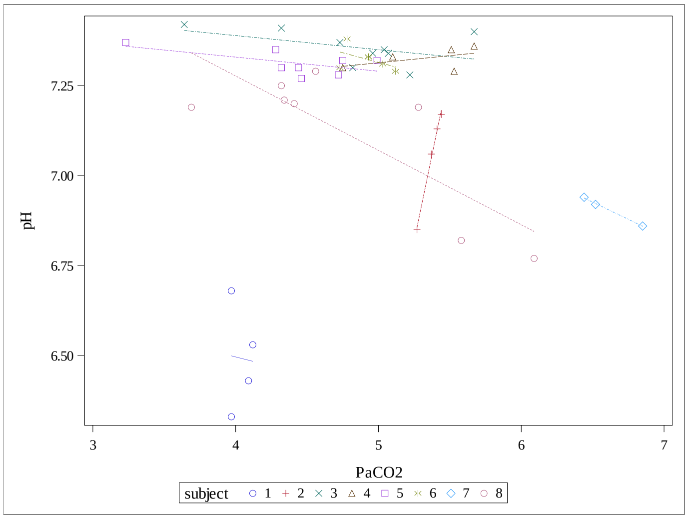

Figure 3. Within-Subject Correlations Between pH and $PaCO_2$

MODEL FIT AND SPECIFICATION

Scenario 1: Equal Within-subject Variances For *x* and *y* And Zero Within-subject Temporal Links Over Time

$$V_i = J_i \otimes \Sigma_B + K_i \otimes \Sigma_W$$

$$with\ \Sigma_B = \text{UN};\ K_i = I_i;\ \Sigma_W = \text{CS}$$

The first model assumes the between-subject variance-covariance Σ_B has an UN structure and the within-subject variance-covariance Σ_W has a CS structure. K_i is a square identity matrix of dimension *mi*, which assumes all correlations between the repeated measures are explained by the between-subject links.

The following SAS statements invoke PROC MIXED to fit a linear mixed-effects model. The METHOD = `ml` option is to specify the maximum likelihood method of estimation for covariance parameters. If no METHOD = options is given, the REML estimation is the default option. The CLASS statement specifies `subject`, `method`, and `time` as classification effects. The MODEL statement specifies the mean model with `y` as the numeric response variable and `method` as the fixed effect. Kenward Roger (`ddfm = kr`) is the method for computing the denominator degrees of freedom for the tests of fixed effects.

The RANDOM statement defines a random slope model of `method`. The SUBJECT = `subject` option instructs PROC MIXED to make the *N* x *N* variance-covariance G matrix for the entire data to be block diagonal, with block corresponding to subject. Each block $G_i$consists of *mi* x *mi* sub-blocks of 2 x 2 between-subject variance-covariance Σ_B, and thus has a dimension of 2*mi* x 2*mi*. TYPE `= un` option defines unstructured matrix for G_i. Unstructured covariance matrix allows different variances for each method and a covariance between them. Inter-independence is assumed across subjects. G and GCORR options request the estimated between-subject variance-covariance G_i and correlations to be printed. V = 7 and VCORR = 7 options demand the estimated variance-covariance V_i and correlations to be printed for the #7 subject.

The REPEATED statement defines a repeated effect of `method`. The SUBJECT = `time(subject)` produces a block-diagonal structure in *N* x *N* variance-covariance R matrix with identical blocks R_i. R_i is a 2*mi* x 2*mi* matrix and is block diagonal of 2 x 2 within-subject variance-covariance Σ_W. TYPE = `cs` defines that the variance-covariance matrix Σ_W has a compound symmetric structure. R requests that blocks of the estimated R matrix, R_i, be displayed. RCORR = `7` produces the correlation matrix corresponding to R_i for subject #7.

The following SAS codes are used to fit the above described linear mixed-effects model:

```
proc mixed data = example method = ml;
   class subject method time;
   model y = method / solution ddfm = kr;
   random method / type = un subject = subject g gcorr v = 7 vcorr = 7;
   repeated method / type = cs subject = time(subject) r rcorr;
run;
```

Selected SAS outputs are present in Tables 3.1-3.5, where labels have been added for clarity.

Estimated R Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.1331	-0.0280
$PaCO_2$	-0.0280	0.1331

Table 3.1. Estimated Within-Subject Variance-Covariance R_i Matrix for Subject #7

Estimated G Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.0435	0.0263
$PaCO_2$	0.0263	0.4711

Table 3.2. Estimated Between-Subject Variance-Covariance G_i Matrix for Subject #7

Estimated V Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.1766	-0.0017	0.0435	0.0263	0.0435	0.0263
Y_1	-0.0017	0.6042	0.0263	0.4711	0.0263	0.4711
X_2	0.0435	0.0263	0.1766	-0.0017	0.0435	0.0263
Y_2	0.0263	0.4711	-0.0017	0.6042	0.0263	0.4711
X_3	0.0435	0.0262	0.0435	0.0263	0.1766	-0.0017
Y_3	0.0263	0.4711	0.0263	0.4711	-0.0017	0.6042

Table 3.3. Estimated Overall Subject Variance-Covariance V_i Matrix for Subject #7

Estimated V Correlation Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	1.0000	-0.0051	0.2463	0.0805	0.2463	0.0805
Y_1	-0.0051	1.0000	0.0805	0.7797	0.0805	0.7797
X_2	0.2463	0.0805	1.0000	-0.0051	0.2463	0.0805
Y_2	0.0805	0.7797	-0.0051	1.0000	0.0805	0.7797
X_3	0.2463	0.0805	0.2463	0.0805	1.0000	-0.0051
Y_3	0.0805	0.7797	0.0805	0.7797	-0.0051	1.0000

Table 3.4. Estimated Correlation between pH and PaCO2 for Subject #7

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.0169	0.2493	7.63	20.13	<.0001
pH	2.1224	0.2561	7.2	8.29	<.0001
$PaCO_2$	0	.	.	.	.

Table 3.5. Solution for Fixed Effects from The Linear Mixed-effects Model

Table 3.1 provides the within-subject variance-covariance matrix $\boldsymbol{R_i}$. Table 3.2 presents the between-subject variance-covariance matrix $\boldsymbol{G_i}$. Table 3.3 shows the estimated overall variance-covariance $\boldsymbol{V_i}$ matrix. Table 3.4 gives the corresponding correlations associated with the $\boldsymbol{V_i}$ matrix. Table 3.5 demonstrate the estimated fixed effects from the linear mixed-effects model. The variances for pH and $PaCO_2$ are 0.1766 and 0.6042 (Table 5). The estimated correlation ρ_{XY} between pH and $PaCO_2$ is -0.0051; the estimated correlation $\rho_{XX'}$ between $x_{ij}, x_{ij'}$ is 0.2563; the estimated correlation $\rho_{YY'}$ between $y_{ij}, y_{ij'}$ is 0.7797; and the estimated correlation $\rho_{XY}\delta$ between $x_{ij}, y_{ij'}$ is 0.0805 (Table 6), for j ≠ j'.

Scenario 2: Unequal Within-subject Variances For *x* and *y* And Zero Within-subject Temporal Links Over Time

$$\boldsymbol{V_i} = \boldsymbol{J_i} \otimes \boldsymbol{\Sigma_B} + \boldsymbol{K_i} \otimes \boldsymbol{\Sigma_W}$$

$$\boldsymbol{with}\ \boldsymbol{\Sigma_B}\ =\ \mathbf{UN};\ \boldsymbol{K_i} = \boldsymbol{I_i};\ \boldsymbol{\Sigma_W} = \mathrm{UN}$$

The second model assumes both the between-subject variance-covariance Σ_B and the within-subject variance-covariance Σ_W have UN structures. K_i is a square identity matrix of dimension *mi*, which assumes all correlations between the repeated measures are explained by the between-subject links.

The REPEATED statement defines a repeated effect of `method`. TYPE = `un` defines that the variance-covariance matrix Σ_W have an unstructured structure.

The following SAS codes are used to fit the above described linear mixed-effects model:

```
proc mixed data = example method = ml;
   class subject method time;
   model y = method / solution ddfm = kr;
   random method / type = un subject = subject g gcorr v = 7 vcorr = 7;
   repeated method / type = un subject = time(subject) r rcorr;
run;
```

Selected SAS outputs are present in Tables 4.1-4.5, where labels have been added for clarity.

Estimated R Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.0116	-0.0276
$PaCO_2$	-0.0276	0.2547

Table 4.1. Estimated Within-Subject Variance-Covariance $\boldsymbol{R_i}$ Matrix for Subject #7

Estimated G Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.0747	0.0252
$PaCO_2$	0.0252	0.4252

Table 4.2. Estimated Between-Subject Variance-Covariance $\boldsymbol{G_i}$ Matrix for Subject #7

Estimated V Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.0862	-0.0024	0.0747	0.0252	0.0747	0.0252
Y_1	-0.0021	0.6799	0.0252	0.4252	0.0252	0.4252
X_2	0.0747	0.0252	0.0862	-0.0024	0.0747	0.0252
Y_2	0.0252	0.4252	-0.0024	0.6799	0.0252	0.4252
X_3	0.0747	0.0252	0.0747	0.0252	0.0862	-0.0024
Y_3	0.0252	0.4252	0.0252	0.4252	-0.0024	0.6799

Table 4.3. Estimated Overall Subject Variance-Covariance $\boldsymbol{V_i}$ Matrix for Subject #7

Estimated V Correlation Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	1.0000	-0.0100	0.8659	0.1042	0.8659	0.1042
Y_1	-0.0100	1.0000	0.1042	0.6254	0.1042	0.6254
X_2	0.8659	0.1042	1.0000	-0.0100	0.8659	0.1042
Y_2	0.1042	0.6254	-0.0100	1.0000	0.1042	0.6254
X_3	0.8659	0.1042	0.8659	0.1042	1.0000	-0.0100
Y_3	0.1042	0.6254	0.1042	0.6254	-0.0100	1.0000

Table 4.4. Estimated Correlation between pH and PaCO2 for Subject #7

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.0082	0.2436	7.25	20.56	<.0001
pH	2.1069	0.2530	7.18	8.33	<.0001
$PaCO_2$	0	.	.	.	.

Table 4.5. Solution for Fixed Effects from The Linear Mixed-effects Model

The variances for pH and $PaCO_2$ are 0.0862 and 0.6799 (Table 4.3). The estimated correlation ρ_{XY} between pH and $PaCO_2$ is -0.0100; the estimated $\rho_{XX'}$ between $x_{ij}, x_{ij'}$ is 0.8659; the estimated $\rho_{YY'}$ between $y_{ij}, y_{ij'}$ is 0.6254; and the estimated $\rho_{XY}\delta$ between $x_{ij}, y_{ij'}$ is 0.1042 (Table 4.4), for j ≠ j'.

Scenario 3: Unequal Within-subject Variances For *x* and *y* And CS Within-subject Temporal Links Over Time

$$\boldsymbol{V_i} = \boldsymbol{J_i} \otimes \boldsymbol{\Sigma_B} + \boldsymbol{K_i} \otimes \boldsymbol{\Sigma_W}$$

$$\boldsymbol{with}\ \boldsymbol{\Sigma_B} = \textbf{UN};\ \boldsymbol{K_i} = \text{CS};\ \boldsymbol{\Sigma_W} = \text{UN}$$

The third model assumes both the between-subject variance-covariance Σ_B and the within-subject variance-covariance Σ_W have UN structures. K_i is a *mi* x *mi* matrix and has a CS structure which assumes that all within-subject correlations between the repeated measures are same.

The REPEATED statement defines a repeated effect of `method` and `time`. The SUBJECT = `subject` produces a block-diagonal structure in *N* x *N* variance-covariance R matrix with identical blocks R_i. R_i is a 2*mi* x 2*mi* matrix and is block diagonal of 2 x 2 within-subject variance-covariance Σ_W.

TYPE = `UN@CS` specify direct (Kronecker) product structures designed for multivariate repeated measures. These structures are constructed by taking the Kronecker product of an unstructured matrix $\boldsymbol{K_i}$ (modeling covariance across the multivariate observations) with an additional covariance matrix $\boldsymbol{\Sigma_W}$ (modeling covariance across time or another factor). To use these structures in the REPEATED statement, you must specify two distinct REPEATED effects, both of which must be included in the CLASS statement. The first effect indicates the multivariate observations, and the second identifies the levels of time or some additional factor. Note that the input data set must still be constructed in "univariate" format; that is, all dependent observations are still listed observation-wise in one single variable. Although this construction provides for general modeling possibilities, it forces you to construct variables indicating both dimensions of the Kronecker product.

R requests that blocks of the estimated R matrix, R_i, be displayed. RCORR produces the correlation matrix corresponding to R_i for subject #7.

The following SAS codes are used to fit the above described linear mixed-effects model:

```
proc mixed data = example method = ml;
   class subject time method;
   model y = method time method*time / solution ddfm = kr;
   random method / type = un subject = subject g gcorr v vcorr = 7;
   repeated method time / type = un@cs subject = subject r rcorr = 7;
run;
```

Selected SAS outputs are present in Tables 5.1-5.5, where labels have been added for clarity.

Estimated R Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.0115	-0.0275	-0.0001	0.0002	-0.0001	0.0002
Y_1	-0.0275	0.2533	0.0002	-0.0014	0.0002	-0.0014
X_2	-0.0001	0.0002	0.0115	-0.0275	-0.0001	0.0002
Y_2	0.0002	-0.0014	-0.0275	0.2533	0.0002	-0.0014
X_3	-0.0001	0.0002	-0.0001	0.0002	0.0115	-0.0275
Y_3	0.0002	-0.0014	0.0002	-0.0014	-0.0275	0.2533

Table 5.1. Estimated Within-Subject Variance-Covariance $\boldsymbol{R_i}$ Matrix for Subject #7

Estimated G Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.0747	0.0251
$PaCO_2$	0.0251	0.4266

Table 5.2. Estimated Between-Subject Variance-Covariance $\boldsymbol{G_i}$ Matrix for Subject #7

Estimated V Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.0862	-0.0024	0.0747	0.0252	0.0747	0.0252
Y_1	-0.0024	0.6799	0.0252	0.4252	0.0252	0.4252
X_2	0.0747	0.0252	0.0862	-0.0024	0.0747	0.0252
Y_2	0.0252	0.4252	-0.0024	0.6799	0.0252	0.4252
X_3	0.0747	0.0252	0.0747	0.0252	0.0862	-0.0024
Y_3	0.0252	0.4252	0.0252	0.4252	-0.0024	0.6799

Table 5.3. Estimated Overall Subject Variance-Covariance $\boldsymbol{V_i}$ Matrix for Subject #7

Estimated V Correlation Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	1.0000	-0.00995	0.8659	0.1042	0.8659	0.1042

Y_1	-0.0100	1.0000	0.1042	0.6254	0.1042	0.6254
X_2	0.8659	0.1042	1.0000	-0.0100	0.8659	0.1042
Y_2	0.1042	0.6254	-0.0100	1.0000	0.1042	0.6254
X_3	0.8659	0.1042	0.8659	0.1042	1.0000	-0.0100
Y_3	0.1042	0.6254	0.1042	0.6254	-0.0100	1.0000

Table 5.4. Estimated Correlation between pH and PaCO2 for Subject #7

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.0082	0.2436	7.25	20.56	<.0001
pH	2.1069	0.2530	7.18	8.33	<.0001
$PaCO_2$	0	.	.	.	.

Table 5.5. Solution for Fixed Effects from The Linear Mixed-effects Model

The variances for pH and $PaCO_2$ are 0.0862 and 0.6799 (Table 5.3). The estimated correlation ρ_{XY} between pH and $PaCO_2$ is -0.0100; the estimated correlation $\rho_{XX'}$ between $x_{ij}, x_{ij'}$ is 0.8659; the estimated correlation $\rho_{YY'}$ between $y_{ij}, y_{ij'}$ is 0.6254; and the estimated correlation $\rho_{XY}\delta$ between $x_{ij}, y_{ij'}$ is 0.1042 (Table 5.4), for j ≠ j'.

Scenario 4: Unequal Within-subject Variances For *x* and *y* And AR(1) Within-subject Temporal Links Over Time

$$\boldsymbol{V_i} = \boldsymbol{J_i} \otimes \boldsymbol{\Sigma_B} + \boldsymbol{K_i} \otimes \boldsymbol{\Sigma_W}$$

$$\boldsymbol{with\ \Sigma_B} = \mathbf{UN};\ \boldsymbol{K_i} = \mathrm{AR}(1);\ \boldsymbol{\Sigma_W} = \mathrm{UN}$$

The fourth model assumes both the between-subject variance-covariance Σ_B and the within-subject variance-covariance Σ_W have UN structures. K_i is a *mi* x *mi* matrix and has an AR(1) structure which assumes that all within-subject correlations between the repeated measures decreases along time.

The REPEATED statement defines a repeated effect of `method`. The SUBJECT = `subject` produces a block-diagonal structure in *N* x *N* variance-covariance R matrix with identical blocks R_i. R_i is a 2*mi* x 2*mi* matrix and is block diagonal of 2 x 2 within-subject variance-covariance Σ_W. TYPE= `UN@AR(1)` specify direct (Kronecker) product structures designed for multivariate repeated measures. These structures are constructed by taking the Kronecker product of an unstructured matrix $\boldsymbol{\Sigma_W}$ (modeling covariance across the multivariate observations) with an additional AR(1) covariance matrix $\boldsymbol{K_i}$ (modeling covariance across time or another factor). R requests that blocks of the estimated R matrix, R_i, be displayed. RCORR produces the correlation matrix corresponding to R_i for subject #7.

The following SAS codes are used to fit the above described linear mixed-effects model:

```
proc mixed data = example method = ml;
   class subject method time;
   model y = method / solution ddfm = kr;
   random method / type = un subject = subject g gcorr v vcorr = 7;
   repeated method time / type = un@ar(1) subject = subject r rcorr = 7;
run;
```

Selected SAS outputs are present in Tables 6.1-6.5, where labels have been added for clarity.

Estimated R Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.0183	-0.0622	0.0135	-0.0460	0.0100	-0.0339
Y_1	-0.0622	0.7697	-0.0460	0.5684	-0.0339	0.4197
X_2	0.0135	-0.0460	0.0183	-0.0622	0.0135	-0.0460
Y_2	-0.0460	0.5684	-0.0622	0.7697	-0.0460	0.5684
X_3	0.0100	-0.0339	0.0135	-0.0460	0.0183	-0.0622
Y_3	-0.0339	0.4197	-0.0460	0.5684	-0.0622	0.7697

Table 6.1. Estimated Within-Subject Variance-Covariance $\boldsymbol{R_i}$ Matrix for Subject #7

Estimated G Matrix for Subject 7		
	pH	$PaCO_2$
pH	0.0690	0.0743
$PaCO_2$	0.0743	

Table 6.2. Estimated Between-Subject Variance-Covariance $\boldsymbol{G_i}$ Matrix for Subject #7

Estimated V Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	0.0873	0.0120	0.0825	0.0283	0.0790	0.0403
Y_1	0.0120	0.7697	0.0283	0.5684	0.0403	0.4197
X_2	0.0825	0.0283	0.0873	0.0120	0.0825	0.0283
Y_2	0.0283	0.5684	0.0120	0.7697	0.0283	0.5684
X_3	0.0790	0.0403	0.0825	0.0283	0.0873	0.0120
Y_3	0.0403	0.4197	0.0283	0.5684	0.0120	0.7697

Table 6.3. Estimated Overall Subject Variance-Covariance $\boldsymbol{V_i}$ Matrix for Subject #7

Estimated V Correlation Matrix for Subject 7						
	X_1	Y_1	X_2	Y_2	X_3	Y_3
X_1	1.0000	0.0464	0.9453	0.1092	0.9048	0.1556
Y_1	0.0464	1.0000	0.1092	0.7385	0.1556	0.5453
X_2	0.9453	0.1092	1.0000	0.0464	0.9453	0.1092
Y_2	0.1092	0.7385	0.0464	1.0000	0.1092	0.7385
X_3	0.9048	0.1556	0.9453	0.1092	1.0000	0.0464
Y_3	0.1556	0.5453	0.1092	0.7385	0.0464	1.0000

Table 6.4. Estimated Correlation between pH and PaCO2 for Subject #7

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.0333	0.2295	8.08	21.93	<.0001
pH	2.0844	0.2293	7.01	9.09	<.0001
$PaCO_2$	0	.	.	.	.

Table 6.5. Solution for Fixed Effects from The Linear Mixed-effects Model

The variances for pH and $PaCO_2$ are 0.0873 and 0.7697 (Table 6.3). The estimated correlation ρ_{XY} between pH and $PaCO_2$ is 0.0464; the estimated correlation $\rho_{XX'}$ between $x_{ij}, x_{ij'}$ is 0.9453; the estimated correlation $\rho_{YY'}$ between $y_{ij}, y_{ij'}$ is 0.7385; and the estimated correlation $\rho_{XY}\delta$ between $x_{ij}, y_{ij'}$ is 0.1092 (Table 6.4), for j ≠ j'.

Scenario 5: Unequal Within-subject Variances For *x* and *y* And UN Within-subject Temporal Links Over Time

$$\boldsymbol{V_i} = \boldsymbol{J_i} \otimes \boldsymbol{\Sigma_B} + \boldsymbol{K_i} \otimes \boldsymbol{\Sigma_W}$$

$$\boldsymbol{with\ \Sigma_B} = \mathbf{UN};\ \boldsymbol{K_i} = \mathrm{UN};\ \boldsymbol{\Sigma_W} = \mathrm{UN}$$

The fifth model assumes both the between-subject variance-covariance Σ_B and the within-subject variance-covariance Σ_W have UN structures. K_i is a *mi* x *mi* matrix and has an UN structure which assumes that all within-subject correlations between the repeated measures differ from each other.

The REPEATED statement defines a repeated effect of `method`. The SUBJECT = `subject` produces a block-diagonal structure in *N* x *N* variance-covariance R matrix with identical blocks R_i. R_i is a 2*mi* x 2*mi* matrix and is block diagonal of 2 x 2 within-subject variance-covariance Σ_W. TYPE= `UN@UN` specify direct (Kronecker) product structures designed for multivariate repeated measures. These structures are constructed by taking the Kronecker product of an unstructured matrix $\boldsymbol{\Sigma_W}$ (modeling covariance across the multivariate observations) with an additional unstructured matrix $\boldsymbol{K_i}$ (modeling covariance across time or another factor). R requests that blocks of the estimated R matrix, R_i, be displayed. RCORR produces the correlation matrix corresponding to R_i for subject #7.

```
proc mixed data = example method = ml;
   class subject time method;
   model y = method time method*time / solution ddfm = kr;
   random method / type = un subject = subject g gcorr v vcorr = 7;
   repeated method time / type = un@un subject = subject r rcorr = 7;
run;
```

The estimation iteration is stopped because of infinite likelihood and thus no result is obtained.

Table 7 shows the summary of the linear mixed-effects model results for scenarios 1 to 4. It indicates that the correct specification of the variance-covariance structure is critically important for an accurate estimation of the correlation between two variables.

Estimate	Scenario 1	Scenario 2	Scenario 3	Scenario 4
μ_X	7.1393	7.1151	7.1151	7.1177
μ_Y	5.0169	5.0082	5.0082	5.0333
$\hat{\boldsymbol{\Sigma}}_B$				
σ_{XB}^2	0.0435	0.0747	0.0747	0.0690
σ_{YB}^2	0.4711	0.4252	0.4266	0.0000
σ_{XYB}	0.0263	0.0252	0.0251	0.0743
ρ_{XYB}	0.1836	0.1415	0.1404	-
$\hat{\boldsymbol{\Sigma}}_W$				
σ_{XW}^2	0.1331	0.0116	0.0115	0.0183
σ_{YW}^2	0.1331	0.2547	0.2533	0.7697
σ_{XYW}	-0.0280	-0.0276	-0.0275	-0.0622
ρ_{XYW}	-0.2100	-0.5092	-0.5092	-0.5250
δ_{XYW}	-	-	-0.0055	0.7385
$\hat{\boldsymbol{\Sigma}}$				
σ_X^2	0.1766	0.0862	0.0862	0.0873
σ_Y^2	0.6042	0.6799	0.6799	0.7697
σ_{XY}	-0.0017	-0.0024	-0.0024	0.0120
$\boldsymbol{\rho_{XY}}$	**-0.0051**	**-0.0100**	**-0.0100**	**0.0464**
$\rho_{XX'}$	0.2463	0.8659	0.8659	0.9453
$\rho_{YY'}$	0.7797	0.6254	0.6254	0.7385
δ_{XY}	-15.7456	-10.4724	-10.4724	2.3529
AIC	121.8	44.4	46.4	24.0
BIC	122.4	45.1	47.2	24.6

Table 7. Estimates from The Linear Mixed-effects Models

CONCLUSION

In this paper, we demonstrate how to use the CCRM (Correlation Coefficient for Repeated Measures) method to calculate correlation coefficient between two variables when repeated observations are available on each of the variables. We first review the development of the CCRM method under the framework of correlation partitioning with the inclusion of temporal links over time. We then illustrate how to apply the CCRM method using the PROC MIXED procedures in SAS to obtain the parameter estimates of interest.

A longitudinal setting is assumed when the two variables are linked over time. Five different scenarios are considered to address the various correlation patterns. SAS example codes as well as examples of hands-on data analysis and outputs are presented. The results indicates that the correct specification of the variance-covariance structure is critically important for an accurate estimation of the correlation between two variables.

REFERENCES

Bland, J.M. and Altman, D.G. 1994. “Correlation, Regression and Repeated Data” *British Medical Journal*, 308, 896.

Bland, J.M. and Altman, D.G. 1995. “Calculating Correlation Coefficients with Repeated Observations: Part 1 – Correlation Within Subjects” *British Medical Journal*, 310, 446.

Lam, Miu, Webb, CA, O’Donnell, DE. 1999. “Correlation between Two Variables in Repeated Measures” *American Statistical Associating, Proceeding of the Biometric Section*, Alexandria, VA: ASA.

Hamlett, Anthony, Ryan Louise, Serrano-Trespalacios Paulina, and Wolfinger, Russ. 2003. “Mixed Models for Assessing Correlation in The Presence of Replication” *Journal of the Air & Waste Management Association*, 53(4):442-450.

Hamlett, Anthony, Ryan, Louise, and Wolfinger, Russ. 2004. “On The Use of PROC MIXED to Estimate Correlation In The Presence of Repeated Measures” *SUGI 29 Proceedings*. Montreal, Canada.

Roy, Anuradha. 2006. “Estimating Correlation Coefficient between Two Variables with Repeated Observations using Mixed Effects Model” *Biometrical Journal*, 48(2):286-301.

Roy, Anuradha, Fuller, Clifton, Rosenthal, David, and Thomas Charles, 2015. “Comparison of Measurement Methods with A Mixed Effects Procedure Accounting for Replicated Evaluations (COM3PARE): Method Comparison Algorithm Implementation for Head and Neck IGRT Positional Verification” *BMC Medical Imaging*, 15(35):1-11.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Qinlei Huang
Merck
267.305.5184
qinlei.huang@merck.com