

PP02: Bivariate Analysis Dataset Structure (ADBIVAR)

Analysis-Ready Dataset Structure for Bivariate Analysis Based on BDS Principles

Michael Pannucci and Toshio Kimura

Abstract

“Analysis-ready” is a key concept of the ADaM-BDS framework whereby a formulation of a where-clause yields a dataset that can be passed to a statistical analysis or modeling procedure with minimal, if any, further manipulation. However, current BDS standards are limited to single-endpoint analysis; that is, an analysis or model that relates one endpoint to a set of factors such as covariates, strata, or other independent variable(s). Consider the case where the analysis is not an endpoint as a function of other factors, but rather an analysis simultaneously involving two distinct endpoints. A common example of this would be a correlation analysis. Within SAS, the existing BDS structure would not be “analysis-ready” when using the CORR procedure. A fair amount of data manipulation, including, but not limited to, Cartesian products and transpositions of data, is required before calling the procedure. While a simpler transposing on visit might be satisfactory for certain situations, the user is then limited to only correlating two endpoints at the same visit. It might be of interest to correlate across visits, thus necessitating a more flexible structure.

In this paper, we examine a natural extension of the BDS framework called the Bivariate Analysis Data Structure (ADBIVAR) that creates two sets of endpoint “triplets” consisting of 1) parameter variable, 2) visit variable, and 3) response variable (AVAL, CHG, PCHG). Hence, the first endpoint triplet would be PARAMCD1, AVISIT(N)1, and RSPVAR1, and the second endpoint triplet will be PARAMCD2, AVISIT(N)2, and RSPVAR2. Therefore, a single observation in ADBIVAR contains the necessary information to correlate “endpoint 1” with “endpoint 2,” thus making ADBIVAR analysis-ready and enabling a call to the CORR procedure without further manipulation.

Analysis Need

Many analyses require two sets of values from a BDS dataset to be placed in the same record for SAS to execute the statistical procedure. Examples include:

Correlation:

- What if we want to correlate a baseline value of one endpoint with a key post-baseline visit of a different endpoint?
- Or correlate the value of one endpoint with the percent change from baseline of another endpoint?
- Or correlate an ADaM variable with any BDS endpoint?

Other Analysis:

- Scatter plot to show the relationship of any two endpoint values
- AUC analysis to evaluate the performance of a continuous outcome threshold to predict a binary outcome
- Cross-tabulation and Kappa analysis of two categorical endpoints

BDS to Analysis:



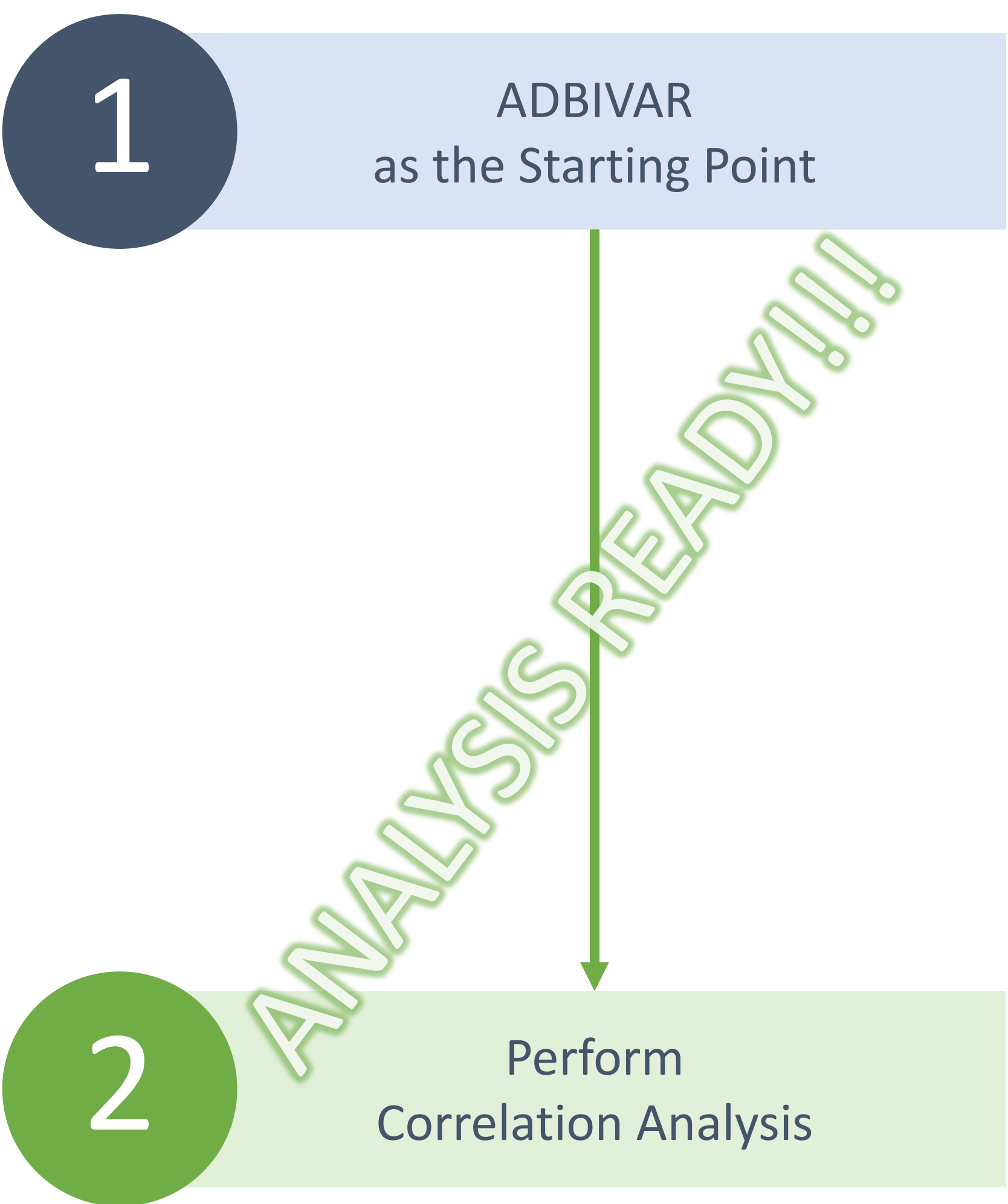
Method

The BDS dataset cannot be used to directly perform a correlation analysis. The required steps are outlined in the graphic on the left. In short, this includes subsetting, renaming and merging. This results in a dataset that has 2 sets of parameter and result information. The proposal is to apply a standard structure to this dataset which we have named ADBIVAR.

ADaM has 2 complete sets of variables to flexibly represent the selection of any 2 BDS endpoint values (or a mixture with ADaM variables). As shown below, PARAM/PARAMCD/PARAMN will be repeated with a suffix of 1 or 2 to represent endpoint 1 or 2. Similar, variables such as AVISIT, AVISITN, RSPVAR, RSPVARN, and RESULT will also be repeated and suffixed accordingly.

ADBIVAR is now analysis-ready. After subsetting with a Where clause, ADBIVAR can be directly used to perform a correlation analysis.

ADBIVAR to Analysis:



ADBIVAR Variables		
Endpoint 1	Endpoint 2	Purpose
PARAM1, PARAMCD1, PARAMN1	PARAM2, PARAMCD2, PARAMN2	Parameter, Parameter code, Parameter sort for each bivariate pair
PARCAT1	PARCAT2	Parameter category for each bivariate pair
AVISIT1, AVISITN1	AVISIT2, AVISITN2	Analysis visit, Analysis visit sort for each bivariate pair
RSPVAR1, RSPVARN1	RSPVAR2, RSPVARN2	Response variable, Response variable sort for each bivariate pair
RESULT1	RESULT2	Analysis value for each bivariate pair

ADBIVAR Structure:		Endpoint 1					Endpoint 2				
		PARAMCD1	PARCAT1	AVISITN1	RSPVAR1	RESULT1	PARAMCD2	PARCAT2	AVISITN2	RSPVAR2	RESULT2
Pivotal-1	P1-10001	VL	BDS	0	BASE	190,000	CD4	BDS	24	AVAL	980
Pivotal-1	P1-10001	VL	BDS	12	AVAL	250	CD4	BDS	24	PCHG	33.7
Pivotal-1	P1-10001	VL	BDS	0	BASE	190,000	ARVDOSE	ADSL	.	ARVDOSE	3

Conclusion

The ADBIVAR structure is a way of replicating key vertical columns from a BDS into triplets of parameter, visit, and response value across two vectors of data. Once these two triplets are formatted in a way such that all data appears in one observation, the analysis-ready requirement for PROC CORR is complete. This structure is also easily generalizable to more than two endpoints.