Journal of Research in Personality 54 (2015) 40–50

Contents lists available at ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

The comprehensive approach to analyzing multivariate constructs

Ryne A. Sherman *, David G. Serfass

Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33433, United States

ARTICLE INFO

Article history:
Available online 20 May 2014

Keywords:
Multivariate constructs
Profile correlations
R
Replicability

ABSTRACT

Many psychological constructs of interest to personality psychologists, such as personality, behavior, and emotions, are made up of many variables. Moreover, similarity metrics, such as self-other agreement, profile similarity, or behavioral consistency, result from calculations conducted across many variables. When analyzed using a comprehensive approach, such *multivariate constructs* present unique analytic challenges. Such challenges are not well addressed in standard graduate statistics textbooks or presently available in standard commercial software. This article introduces the 'multicon' package, freely available in the R statistical package, designed to aid researchers interested in taking a comprehensive approach to analyzing multivariate constructs. Realistic examples from personality psychology are provided to demonstrate the utility of this package.

1. Introduction

Is personality related to behavior? How do extraverts behave differently from introverts? How well do two people agree about what someone else's personality is like? How accurately can we judge someone else's personality? How similar/consistent are people or situations? Personality scientists are often concerned with these sorts of questions and many more like them. However, answering questions such as these can be quite complicated. To see why, compare these questions to another question: What is the relationship between a person's height and weight? A key difference is that the constructs of interest in the first set of questions are multivariate, while the constructs in the latter question are not. *Multivariate constructs*, as the name implies, refer to psychological constructs that consist of many psychological variables.[1] Many constructs of interest to personality psychologists are multivariate in nature: personality, behavior, emotions, motives, situations, etc.

The difficulty with multivariate constructs is that they make answering questions like those posed in at the outset challenging. For example, answering the question about the relationship between personality and behavior requires, at minimum, some definition of what is meant by "personality" and "behavior." Depending on one's particular perspective, the multivariate construct of *personality* might include thousands of traits (Allport & Odbert, 1936), one-hundred (Block, 1961), or merely a handful (i.e., 5; McCrae & Costa, 2008). Regardless, most personality scientists recognize that personality is a multivariate construct. Behavior is also a multivariate construct although arguably psychologists have put less effort into taxonomizing behavior than personality (Furr, 2009).

There are roughly two strategies psychologists have used to deal with the problem of multivariate constructs.[2] The first strategy reduces the construct(s) of interest to a smaller number (e.g., 1–6) of more mentally tractable, often empirically derived, essential variables. We refer to this strategy as the *essential approach*. For example, instead of "personality" (broadly construed) one might focus on just a *single* trait (e.g., extraversion) or a subset of broad traits (e.g., the Big 5). Likewise, instead of "behavior" (broadly construed) one might focus on just a *single* behavior (e.g., talkativeness) or on a subset of broad behaviors (e.g., interpersonal behaviors from the Interpersonal Circumplex).

The second strategy for dealing with the problem of multivariate constructs tries to avoid data reduction as much as possible preferring to comprehensively assess and analyze the many relationships between the constructs of interest. We refer to this strategy as the *comprehensive approach* (Sherman & Wood, 2014). A researcher employing this approach may use measures designed such that each item represents a distinct characteristic such as the California Adult Q-set (CAQ: Block, 1961) or the Inventory for Individual Differences in the Lexicon (IIDL: Wood, Nye, & Saucier, 2010). Alternatively, a comprehensive approach may even employ

* Corresponding author.
E-mail address: rsherm13@fau.edu (R.A. Sherman).

[1] We are grateful to Mike Furr for suggesting the name *multivariate constructs*.

[2] We say "roughly" because Funder's (2013) *single trait*, *essential trait*, and *many trait* approaches to personality correspond to the strategies we identify. We do not differentiate between Funder's single trait and essential trait approaches here for the sake of simplicity.

http://dx.doi.org/10.1016/j.jrp.2014.05.002
0092-6566/

measures designed to assess *essential* variables (e.g., the NEO PI-R: Costa & McCrae, 1992; the Big Five Inventory: John & Srivastava, 1999; the HEXACO-PI-R: Lee & Ashton, 2004), but treat each item as if it were to be analyzed separately (cf. Biesanz, 2010; Biesanz & Human, 2010; Human & Biesanz, 2011a, 2011b).

There are strengths and weaknesses to both approaches. The essential approach reduces complex multivariate constructs such as personality and behavior into mentally tractable subsets. This makes the research conceptually easier to transmit to other scientists and beyond. The comprehensive approach, on the other hand, can be mentally taxing (i.e., who wants to look at a correlation matrix with $100 \times 67 = 6700$ unique elements?; see Section 1.1). An additional advantage of the essential approach is that it can drastically reduce the number of variables analyzed resulting in lower Type I error rates. The comprehensive approach often involves computing a large number of correlations and risks identifying noise as signal. However, the essential approach may miss or obscure associations between the constructs of interest (cf. Brown & Sherman, 2014; Fast & Funder, 2008; Hirsh, DeYoung, Xu, & Peterson, 2010). The comprehensive approach is less likely to miss or obscure such associations. Lastly, both the essential and comprehensive approaches can be used to answer questions about agreement, similarity, or consistency at the nomethetic (e.g., item) level. However, comprehensive approaches—which include more variables—may be superior for addressing these questions at the ideographic (e.g., person, profile) level because the increased number of variables increases the reliability of such profiles.

A perhaps less-well recognized difference between the essential and comprehensive approaches is that the statistical tools for conducting analyses from an essential approach are well-described in graduate statistics textbooks, widely available in standard commercial software (e.g., SAS, SPSS, Excel), and easy to implement. The comprehensive approach, on the other hand, comes with a unique set of problems (e.g., how to handle so many variables, how to appropriately test for profile similarity) requiring different data analytic methods. Such methods are not (a) well-described in textbooks, (b) widely available in standard commercial software, or (c) easy to implement.

This article introduces the 'multicon' package—an R package offering functions designed to deal with the problems inherent with the comprehensive approach for handling multivariate constructs (Sherman, 2014). In this article, we provide examples of realistic questions a personality scientist may encounter and show how a researcher using a comprehensive approach might use the functions available in the 'multicon' package to address these questions. Table 1 provides a summary of the types of questions we address in this article along with the functions from the 'multicon' package used to address them. All datasets used in these examples are built into the 'multicon' package making it easy to follow along.[3] Although we refer to differences between the essential and comprehensive approaches to handling multivariate constructs, this article is not meant to create, or resolve, a conflict between these two approaches. Indeed, as noted previously, both approaches have strengths and weaknesses. As such, this article will primarily focus on analytic issues involved in using a comprehensive approach and describe the tools provided by the 'multicon' package to help resolve them.

1.1. Are these two multivariate constructs related?

We began by asking what appears to be a simple question: Is personality related to behavior? Let us say that we have measured personality with the 100-item CAQ (Block, 1961) and behavior with the 67-item Riverside Behavioral Q-sort (RBQ: Funder, Furr, & Colvin, 2000; Furr, Wagerman, & Funder, 2010). The essential approach to this question would be to first, for both personality and behavior, reduce the number of items measured to some essential subset. Such subsets could be derived empirically (e.g., factor analysis, principal components) or theoretically (e.g., the interpersonal circumplex; see Markey, Funder, & Ozer, 2003). The second step using the essential approach would then be to examine the associations (correlations) between the resultant subsets of variables. Almost all software packages, commercial or otherwise, are designed to make such analyses easy and convenient.

A comprehensive approach this question though would aim to analyze the full set of correlations between all 100 personality items and the 67 behaviors. Calculating such a correlation matrix is usually quite easy in just about any statistical package. However, as previously noted, perusing through a matrix of 6700 correlations will likely prove mentally intractable. Thus, an alternative method for quantifying the degree of relationship between personality and behavior is needed. One method is to count the total number of statistically significant correlations in the matrix (cf. Block, 1960). Another is to determine if the average magnitude amongst the 6700 correlations is larger than one would expect if the constructs were not related (Sherman & Funder, 2009). Following Sherman and Funder (2009), a randomization test can be used to do both of these simultaneously. The test randomly reassigns CAQ profiles to RBQ profiles, creating a pseudo dataset, and calculates both the total number of statistically significant correlations and the average absolute *r* amongst the 6700 correlations in this pseudo dataset. To better illustrate this process, imagine picking up each subject's CAQ profile (keeping all 100 scores intact) and randomly reassigning this profile to a subject's RBQ profile. In doing so, one is simulating a random relationship between personality and behavior, while maintaining the dependencies (covariation) within the multivariate constructs. Next, one calculates the 100×67 correlation matrix on this pseudo dataset and records the number of statistically significant correlations and the average absolute *r* of this correlation matrix. These numbers represent simulated values under a model of a random relationship between personality and behavior. Repeating this procedure many times allows for the formation of a sampling distribution, to which we can compare the observed results from the original dataset. Calculating the proportion of simulated values greater than or equal to the observed values (for the number statistically significant and the average absolute *r* respectively) yields a *p*-value indicating the probability of obtaining the originally observed results under chance.

Conducting such an analysis using standard commercial software is either not possible or would require an arduous amount of programming. The *rand.test* function in the 'multicon' package conducts such an analysis. In this example, we use the *rand.test* function to determine whether personality (as measured by the CAQ) has an overall relationship with behavior (as measured by the RBQ).

```
install.packages('multicon') # Only if this is the
  first time using this package
library(multicon)# Load the mulitcon package
data(caq)# Loading the CAQ dataset
data(beh.comp) # Loading the behavior dataset
rand.test(caq, beh.comp, sims=10000) # The
  analysis; could take a minute or so
```

It should be noted that because the sims argument is set to 10,000, which is ten times more than the default value, this analysis may take 30 s or more. The output from this analysis is a list

[3] More information about this dataset can be found in Sherman, Nave, and Funder (2010).