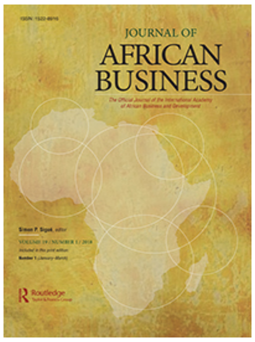

Journal of African Business

McQuitty, Shaun. Os propósitos dos métodos de análise de dados multivariados: um comentário aplicado, Journal of African Business, 19:1, 124-142, 2018.
DOI:10.1080/15228916.2017.1374816
https://www-tandfonline-com.ez43.periodicos.capes.gov.br/doi/epdf/10.1080/15228916.2017.1374816?needAccess=true
OU https://www-tandfonline-com.ez43.periodicos.capes.gov.br/doi/full/10.1080/15228916.2017.1374816?src=recsys#d1e120

ISSN: 1522-8916 (Print) 1522-9076 (Online) Journal homepage: www.tandfonline.com/journals/wjab20

The Purposes of Multivariate Data Analysis Methods: an Applied Commentary

Shaun McQuitty

JOURNAL OF AFRICAN BUSINESS, 2018
VOL. 19, NO. 1, 124–142
https://doi.org/10.1080/15228916.2017.1374816

The Purposes of Multivariate Data Analysis Methods: an Applied Commentary

Shaun McQuitty

Faculty of Business, Athabasca University, Athabasca, Alberta, Canada

ABSTRACT
Unlike univariate and bivariate statistical methods, multivariate statistical methods are capable of analyzing more than one relationship at a time. There are a number of multivariate data analysis methods, each with its own purpose. For nonexperts in statistical analysis, it can be daunting to determine what method is appropriate for a given application. The goal of this commentary is to introduce the multivariate data analysis methods in practical terms that do not require a strong statistical background. The emphasis is describing the purpose of each method, so that readers can choose the methods appropriate for their research questions. The use of multivariate statistical methods in the *Journal of African Business* also is explored.

KEYWORDS
Commentary; business research; quantitative research methods; multivariate statistical analysis; *Journal of African Business* statistical applications

1. Introduction

Statistical methods are used when researchers have quantitative or numerical data that require analysis; the purpose of such analysis is to interpret and summarize the data. Univariate analysis involves the analysis of a single variable, and univariate statistics are descriptive – like a mean or standard deviation – because they describe the data. Bivariate analysis involves two variables, and typical bivariate statistics gauge the strength of the relationship or association between the two variables, e.g., a correlation. Multivariate analysis involves more than two variables, and there is a great variety of multivariate statistical methods that can be used to assess different forms of relationships among variables.

The objective of this commentary is to verbally describe the various multivariate statistical methods, with a focus on their purposes rather than their mathematical foundations. With knowledge of their purposes, researchers may understand the types of research questions that can be addressed by each of the multivariate statistical methods (Osuagwu, 2002). Because research questions, data collection, and data analysis must harmonize with one another, it is essential that they all are considered prior to developing a research design. However, the typical questions asked by researchers who do not know the various statistical methods well are "What method should I use to

CONTACT Shaun McQuitty shaunm@athabascau.ca Faculty of Business, Athabasca University, 1 University Drive, Athabasca, AB, T9S 3A3, Canada

answer my research question?" or "What is the purpose of this statistical method – what does it do?" This commentary answers those questions.

It may appear unreasonable to simplify a difficult topic like multivariate statistical analysis to such a degree. However, the reason we focus on verbal descriptions of the statistical methods' purposes is to encourage their application. Descriptions of multivariate statistical methods in textbooks and published articles tend to be complicated and assume significant background knowledge of mathematics, research methods, and basic statistical methods. Such a presentation requires a considerable amount of learning before researchers can perform statistical analysis.

By concentrating on the purposes of the multivariate statistical methods rather than their mathematical formulations, we assume that readers who do not have strong mathematical and statistical backgrounds can read this commentary and derive an understanding of when each method is appropriate. The intuitive explanations use no mathematics, and should encourage people without a strong background to choose a method and attempt statistical analysis. Further reading will be necessary to use the methods correctly, but method selection is a critical step when conducting quantitative business research.

2. An introduction to multivariate statistics

Multivariate statistics can be described in various ways. For example, univariate and bivariate analyses are special cases of multivariate statistical methods, which are the general case that permits the simultaneous analysis of multiple dependent and independent variables (Tabachnick & Fidell, 2013). Multivariate methods also "simultaneously analyze multiple measurements on individuals or objects under investigation" (Hair, Black, Babin, & Anderson, 2010, p. 4), and thereby encourage the use of multi-item measures of latent constructs to improve measurement reliability and validity (Churchill, 1979; Nunnally & Bernstein, 1994). Multivariate statistics are much more powerful than univariate and bivariate statistics due to their ability to model different forms of relationships among variables. Each multivariate method can evaluate a type of relationship that reflects a particular form of research question, and it is these purposes on which the discussion of different methods is focused.

In a popular textbook that uses business oriented examples, Hair et al. (2010) offer a classification of the multivariate statistical methods (their figure 1–1, pp. 12–13) that is based on: (1) whether variables can be categorized as independent or dependent; (2) the number of dependent variables that can be included in a single statistical model and analysis; and (3) whether the variables are metric (use interval or ratio level measures) or nonmetric (use nominal or ordinal level measures). Tabachnick and Fidell (2013) offer a similar classification using a decision tree that is based on: (1) the nature of the research question; (2) the number of dependent and independent variables; and (3) the number of covariates (their table 2, pp. 29–31). Both of these approaches are representative of rules that can be used to help choose from among the different multivariate statistical methods, but an even more practical approach is to understand the purpose of each method and what kinds of research questions they can address. Providing a useful verbal (non mathematical) explanation of these purposes is the goal of this commentary, and should make multivariate statistical methods more accessible to researchers who are unfamiliar with their use.

The general purpose of all statistical methods – univariate, bivariate, or multivariate – is to summarize, reduce, and interpret raw data. Each of the many different multivariate statistics also has its own specific purpose, and some of the methods have specialized versions that can modify the purpose somewhat. We focus on the primary version of each method, and attempt to identify important variations if their purposes are different from the primary version. Note that the discussion focuses on the methods' purposes, and not their application. A much longer and more detailed discussion is required to explain how to use statistical methods but, having selected a method as appropriate, researchers can then read additional material to gain the expertise necessary to use that method.

Note also that the formal mathematical description of the statistical models may or may not describe the practical purpose of the method. One example is conjoint analysis, which is related mathematically to regression models but, at a practical level, conjoint is similar to ANOVA models because they both compare means. Differences between the practical purpose (what is the statistical method used for?) and the mathematical foundation (how is the statistical model estimated?) can readily be explained, but only if one understands the mathematics underlying the method. Another example is regression and ANOVA models, which both concern variations from the mean at the mathematical level but have different purposes. It is these purposes on which we concentrate the discussion, and not the mathematical foundations.

3. The statistical methods and their purposes

In this section, we introduce the popular multivariate statistical methods and provide a brief description of each method with an explanation of its purpose. Examples of articles that apply the various method applications are provided and drawn mostly from the *Journal of African Business* (not all methods have been used in the journal). In an attempt to organize the descriptions, and to assist researchers attempting to choose among methods, we group the methods into three categories based on their purposes: (1) the correlational methods that focus on explaining observed variables' variances or covariances through their associations with one another, including the canonical correlation, regression analysis, principal component analysis, factor analysis, and structural equation modeling; (2) the difference methods that evaluate variables on the basis of their differences across groups, including the ANOVA family, logistic regression, discriminant analysis, and conjoint analysis; and (3) the distance or similarity methods that locate items of interest relative to one another, and represented by cluster analysis, multidimensional scaling, and correspondence analysis. These categories are not necessarily mutually exclusive, because some methods can have purposes that fall in more than one category (e.g., structural equation models typically focus on explaining covariance and testing theory, but also can evaluate differences in construct means). However, the descriptions emphasize the primary purposes but also explain how a method can have alternative purposes. The brief descriptions of the multivariate methods are given in Table 1, which lists the various methods and summarizes their purposes.

References for examples of published articles are provided for each method to demonstrate how the methods can address certain types of research questions. Where

Table 1. Summary of Multivariate Statistical Method Purposes.

Multivariate Method	Purpose
Canonical Correlation	Assesses the overall strength of relationship (correlation) between two sets of variables.
Multiple Regression	Used when a researcher wants to explain the variance (the variable behavior) of a dependent variable. There can be anywhere from one to many independent (explanatory) variables.
Principal Component Analysis (PCA)	An alternative method of factor analysis that groups variables into components (rather than factors), but the focus is on explaining variables' variances and not their covariances (as in factor analysis).
Factor Analysis (FA)	Purpose is to group variables/items into factors on the basis of their correlation to one another (i.e., purpose is to explain items' covariance). Often used when developing multi-item measurement scales. A related method is principal components analysis, which focuses on creating components (like factors) that explain item variance.
Structural Equation Modeling (SEM)	A method for testing theories about how constructs relate to one another. Also frequently used for evaluating multi-item measurement scales. It combines factor analysis with regression analysis, because each regression variable has multiple measures (like a factor).
Analysis of Variance (ANOVA), Multivariate Analysis of Variance (MANOVA), and MANCOVA	ANOVA is used to compare group means on a metric dependent variable. The groups are created by categorical factors that serve as independent variables. With MANOVA, there are two or more metric dependent variables. MANCOVA is MANOVA with independent covariates that are analyzed in a fashion similar to regression analysis.
Logistic Regression	A special case of regression whose purpose is to divide the object of interest (e.g., respondents) into two groups based on scores from independent variables. Often used with probabilities because logistic regression limits predictions to a 0–1 range.
Discriminant Analysis	The purpose is to place respondents into two or more groups by finding the variables that best differentiate between the groups. In other words, what variables have the biggest mean differences across two or more groups?
Conjoint Analysis	Used to determine the relative importance of different levels of different characteristics, such as a study of a product's attributes. It is an extension of the MANOVA family that compares means across groups. A big difference is that conjoint analysis also permits estimation of individual-level coefficients, rather than group-level only.
Cluster Analysis grupo de Famílias	The purpose is to group similar objects. A typical use of cluster analysis is to group people (into clusters) on the basis of the similarity of their answers to questionnaire items.
Multidimensional Scaling (MDS) and Correspondence Analysis (CA)	The basic purpose of these methods is to infer what characteristics respondents use when comparing alternatives (e.g., jobs or products). The methods also estimate the relative importance of characteristics and how the alternatives compare on these characteristics (this information can be used to create perceptual maps). Estimates can be individual- or group-level. MDS uses similarity or preference data, whereas CA uses cross-table count data.

possible, we selected articles from the *Journal of African Business* to serve as the examples. Note that some of the sample articles use more than one multivariate statistical technique, which shows that statistical methods can be combined to address different research questions from the same study or article. An example is using principal component analysis to obtain component scores, which are then subsequently used in a regression model. Another example is using factor analysis to create and evaluate multi-item scales, and then using structural equation modeling to test a theory about how measured constructs are related.

In his review of articles published in the first 7 years of the *Journal of African Business* (2000–2006), Zoogah (2008) finds that 48% (46/96) are quantitative and 52% are qualitative, so quantitative analyses are well supported by *Journal of African Business* authors. Table 2 lists the same multivariate methods as Table 1, but provides a frequency count of the number of times each method is mentioned in *Journal of African Business* articles available in the ABI/Inform Global database thus far (August 2017, at the time of writing). The count may not perfectly reflect the actual use of methods in articles (e.g., a method could be mentioned but not used, or editorials may describe analysis methods), but it does reflect the relative use and interest in the individual statistical methods. Nevertheless, it can be seen that some multivariate methods are used quite frequently, with factor analysis the most popular method (48 times, although some of these actually are principal component analysis), closely followed by regression (46 times), and then structural equation modeling (16 times), ANOVA (14 times), principal component analysis (14 times), and logistic regression (11 times) used most often. Also interesting is that several of the methods either are little used or do not appear to have been applied at all in published *Journal of African Business* articles, including the canonical correlation, MANCOVA, and correspondence analysis (zero times each), multidimensional scaling (once), conjoint analysis and MANOVA (twice each), and cluster analysis and discriminant analysis (three times each). Knowledge of the purpose of each statistical method may encourage their use in future articles.

3.1. *The correlational methods*

The descriptions begin with the correlational methods that, overall, are the most frequently used methods in business research including articles published in the *Journal of African Business* (Table 2). This category focuses on the correlation or covariance between variables, and includes the canonical correlation, regression analysis, factor analysis, principal component analysis, and structural equation modeling.

Table 2. Frequency of Multivariate Statistical Method Use in *Journal of African Business* Articles.

Multivariate Method	Search Term(s)	Frequency of Use
Canonical Correlation	canonical correlation	0
Multiple Regression	regression analysis	46
Principal Component Analysis (PCA)	principal component	14
Factor Analysis (FA)	factor analysis	48
Structural Equation Modeling (SEM)	structural equation	16
Analysis of Variance (ANOVA), Multivariate Analysis of Variance (MANOVA), and MANCOVA	analysis of variance (ANOVA) (MANOVA, MANCOVA)	13 (14) (2, 0)
Logistic Regression	logistic regression	11
Discriminant Analysis	discriminant analysis	3
Conjoint Analysis	conjoint analysis	2
Cluster Analysis	cluster analysis	3
Multidimensional Scaling (MDS)	multidimensional scaling	1
Correspondence Analysis (CA)	correspondence analysis	0

3.1.1. *Canonical correlation*

We begin with the canonical correlation because – conceptually – the canonical correlation underlies all other correlation-based multivariate statistics, including regression, factor analysis, and structural equation modeling. A canonical correlation assesses the overall strength of association between two sets of variables. More specifically, the canonical correlation uses the variance shared by a set of variables to measure the degree of association between two sets of variables. However, if there is little shared variance in a set of variables, then the canonical correlation with another set of variables may not explain much variance in any of the original set of variables.

A typical example of the use of the canonical correlation is evaluating the overall association between any two multi-item constructs (the relationship between job satisfaction and job turnover intentions is a specific example). Multivariate statistical methods often allow the use of multiple measures (observed variables or items) of the same construct (the underlying phenomenon being measured) to improve measurement reliability and validity. Using multiple measures of the same construct is a foundation of a measurement approach known as Classical Test Theory (see, e.g., Churchill (1979) for an introduction to the ideas underlying Classical Test Theory).

The canonical correlation is less frequently used today than in the past. A large part of the explanation for the diminished use of the canonical correlation is that alternative methods are more specialized, and structural equation models also can estimate the overall correlation between two sets of variables. As evidence of its diminished use over time, Hair, Anderson, Tatham, and Black (1998) has a chapter on the canonical correlation but subsequent editions of the same multivariate textbook do not, and the method has not been applied in the *Journal of African Business* (see Table 2). Nonetheless, by relating two groups of variables, the canonical correlation represents an important model form. It still is possible that some research questions are best evaluated by the method, and there exist articles that demonstrate its use, e.g., Chen, Chen, and Okumus (2013) and Rothmann and Rothmann (2010).

3.1.2. *Multiple regression*

Regression analysis is a popular statistical method that is used when a researcher wants to explain the variance (the behavior) of a single dependent variable. Examples of dependent variables that a researcher might like to explain are variables such as a company's sales figures each month, or a stock/share price, or reasons underlying employee turnover. Each of these dependent variables likely exhibits variation over time, and regression analysis allows researchers to study the effect of one or more additional variables (called explanatory or independent variables) that might explain the behavior or variation of the dependent variable. The independent variables are evaluated in a regression analysis in terms of their ability to explain the dependent variable's behavior.

There are many forms of regression models, and they can range from just one or two independent variables to hundreds of variables. Relatively complicated models can be created by using multiple regression equations with different dependent variables, and where a dependent variable in one equation is an independent variable in another equation (making the model truly multivariate). Independent variables can take other forms; a popular example is time series models that are created by including lagged variables that reflect information from past time periods, including lagged dependent variables. However, the purpose always is to explain a dependent variable's variance, and the ability of a model to explain variance suggests that – to some degree – dependent variable values can be explained and predicted by the independent variables in the model. The proportion of the variance explained is measured by the R^2 statistic (or coefficient of determination), and the parameters associated with each independent variable assess the degree to which they are related to the dependent variable.

Regression is used in many different kinds of applications, and is a fundamental and popular method in business research, as reflected by its frequent use in the *Journal of African Business* (45 times, Table 2). Examples of regression applications can be found in Musila (2013), Osinubi and Amaghionyeodiwe (2003), and Sigué and Bonsu (2012).

3.1.3. *Factor analysis*

Factor analysis is a popular multivariate statistical method that primarily is used to develop and evaluate multi-item measurement instruments; this is the most frequently used method in *Journal of African Business* articles (48 times, Table 2). The main purpose of factor analysis is to group objects into factors on the basis of the objects' correlations with each other. Highly correlated objects can form well defined and easy to interpret factors, whereas objects that are relatively uncorrelated with the others may not be included in a meaningful or interpretable factor. The objects can take the form of variables (typically items in a questionnaire), respondents (e.g., people or organizations), or occasions (Cattell, 1952). The most frequently used form of factor analysis by far is R-type, which groups questionnaire items with numerical responses. Examples of alternative forms of factor analysis are grouping respondents based on their responses to items, or grouping items on the basis of occasions. However, these forms rarely are used, and cluster analysis is the statistical method that tends to be used for grouping respondents.

When a large number of items are used to measure an underlying construct, the factors that result from a factor analysis can be used to determine whether there are different dimensions of the same construct. If the factor analysis groups all items into a single factor, then the factor has just one dimension (is unidimensional). Alternatively, if the factor analysis produces more than one factor, then each factor can be interpreted to understand the dimensions underlying the construct, based on the items associated with each factor. An example is the service quality construct that in one popular measurement scale (SERVQUAL, Parasuraman, Ziethaml, & Berry, 1988) is assumed to have five dimensions, including reliability, responsiveness, assurance, tangibles, and empathy.

comparando com meu projeto poderia ser: clima, relevo, cobertura e uso do solo, unidade de conservação, fitofisionomia, outros.

The most common application of factor analysis is to evaluate questionnaire items to either develop a multi-item scale or to determine whether an existing scale works as expected; Albinsson, Yasanthi, & Sautter (2016) and Mori (2014) are examples. Factor analysis is described as exploratory when evaluating a number of items for inclusion in a multi-item scale (see, e.g., Churchill (1979) for an explanation of the scale development process), and confirmatory when evaluating an existing scale to confirm that items are correlating as expected (items within a factor should load more strongly than

Iramuteq

items across factors). The analysis is the same whether it is confirmatory or exploratory, but the interpretation is different.

Another application of factor analysis is to create factor scores, which are weighted averages of the items comprising a factor. Because the set of items within each factor is summarized by a single numerical figure, factor scores reduce the number of variables from the original number of items to the number of factors, and the new variables (the factor scores) can be used in subsequent analysis with other statistical methods. Note that if the subsequent analysis involves regression models (explaining variance), then scores derived from principal component analysis – known as component scores – likely are more appropriate than factor scores because the component scores retain more of the original items' variances.

3.1.4. *Principal component analysis*

A method strongly related to factor analysis is principal components analysis (or PCA), which is an alternative way of estimating the factor analysis model. By grouping objects (typically questionnaire items) with the greatest correlations, factor analysis attempts to explain the covariance between these items. Principal components analysis also groups items, but the purpose is slightly different because principal components analysis mathematically rearranges items into components on the basis of their variance. In other words, components explain item variances, and not their covariances as in factor analysis. Components are analogous to factors in factor analysis, and can be used to create new variables called component scores; each component score reflects the measured items comprising the component. The purpose of creating the component scores is to explain as much of the original variables' (items) variance as possible with the least number of components. Biboum and Sigué (2014), Gagoitseope and Pansiri (2012), and Kamukama (2013) provide examples of principal components analysis applications.

é semelhante à ponderação/peso?

The ability to create component scores that capture as much of the original variables' variance as possible may be the best use of principal components analysis, because doing so can reduce the number of variables considerably (this benefit was particularly important before computers became so powerful). Component scores can be used in subsequent analysis, but typically are used in regression models where explaining variance is the purpose because the component scores retain as much of the original variables' variance as possible.

It is important to note that factor analysis and principle component analysis often are confused with one another, perhaps because statistical software doesn't always fully differentiate between the two methods. For example, the default estimation method for factor analysis in SPSS is principal component estimation. However, an estimation method like maximum likelihood estimation should be used for factor analysis when explaining covariance is the goal. In practice, the difference is somewhat academic because the solutions obtained by factor analysis with maximum likelihood estimation and with principal components estimation typically are very similar.

3.1.5. *Structural equation modeling*

Despite not being used as frequently as other methods in *Journal of African Business* articles (Table 2), structural equation modeling (SEM) may be the most used multivariate statistical method in the social sciences. It is popular in business research, particularly in the management and marketing areas, and is used extensively in most social science disciplines. SEM also is experiencing greater use in the natural sciences because it is capable of modeling constructs that cannot be measured directly, such as in biology (e.g., Grace, 2006; Pugesek, Tomer, & von Eye, 2003) or medicine (e.g., Batista-Foguet, Coenders, & Ferragud, 2001).

SEM arose from a statistical method known as path analysis and the desire to use multi-item scales for constructs to improve their measurement reliability and validity. The notion of using multiple items to improve measurement follows from Classical Test Theory, which holds that reliability is a positive function of the number of measures in a scale and their degree of correlation (e.g., Churchill, 1979; Nunnally & Bernstein, 1994).

SEM's purpose is to explain the covariance among a set of variables whose relationships are described in advance by the researcher. Put another way, "SEM typically is for testing a theory about the relationships among a group of constructs that are measured with multi-item scales" (McQuitty & Wolf, 2013, p. 68). Such theory can be illustrated with what is known informally as a boxes and arrows diagram. For example, theory could be found to support a model where organizational support, income and benefits, and the work environment all affect job satisfaction, which in turn affects employee turnover intentions and productivity. SEM estimates the expected relationships among the constructs and allows multi-item measures of each construct. Thus, SEM requires that researchers specify their model using a system of equations that permits constructs to be either or both independent and dependent variables. Consequently, Hair et al. (2010) note that "SEM estimates a series of separate, but interdependent, multiple regression equations simultaneously by specifying the structural model used by the statistical program" (p. 617). Examples of SEM applications include Blankson, Mbah, and Owusu-Frempong (2009), Dadzie, Winston, and Dadzie (2012), and Makanyeza and du Toit (2016); but examples are easy to find due to the popularity of the method.

Both SEM and factor analysis have the same purpose – explaining covariance among variables – but SEM goes beyond factor analysis by using the factors as variables in a system of regression models. In this way, SEM is related to both factor analysis and regression. SEM also can be used for purposes other than explaining covariance and theory testing. Like factor analysis, it can be used to assess the measurement properties of a multi-item scale. Reliability and convergent validity are reflected by the size of the parameters between an item and the construct on which it loads, and discriminant validity by not having significant cross loadings (relationships between items and constructs other than the construct on which an item is expected to load). Unidimensionality is implied by convergent and discriminant validity, and is necessary for estimating measurement reliability. SEM can be used to determine the ability of independent constructs to explain the behavior (variance) of dependent constructs; this effectively is regression but with multi-item constructs. With various forms of multi-group analysis, SEMs also can be used to compare groups on the basis of latent means or to model growth curves. Competing theories and models can be compared with SEMs by making changes to the specified relationships, and then comparing the models' goodness-of-fit statistics. With the ability to perform all these types of tests, it is unsurprising that SEM has become so popular in business research.

3.2. *The difference methods*

Broadly speaking, the second category of multivariate statistical methods uses differences in means (mathematical averages) to compare groups. The groups can be defined in many ways, such as by experimental conditions or factors, or by time periods (e.g., before and after), or on the basis of responses to items on a questionnaire (e.g., high and low groups). The statistical methods in this category include the ANOVA family, logistic regression, and discriminant analysis. Conjoint analysis is another method in the category that compares means, but typically across attributes rather than groups. A description of each method and its purpose follows.

3.2.1. *Analysis of variance (ANOVA, MANOVA, MANCOVA)*

Analysis of variance (ANOVA) is an important and much applied statistical method that is used to compare the means of a single variable across groups. The groups are determined on the basis of factors, which are categorical independent variables that can affect a continuous dependent variable that is assumed to vary across the different factor levels (i.e., groups or categories). ANOVA often is used in an experimental setting where, for example, respondents are randomly assigned into groups who experience different conditions and then are measured on a dependent variable. The mean score of the dependent variable can then be compared across the groups to determine if the experiment's manipulation (assignment to groups) had an effect. Gagoitseope and Pansiri (2012) and Winston, Dadzie, and Dadzie (2009) offer examples of ANOVA applications.

If there is more than one continuous dependent variable, then an ANOVA becomes a MANOVA (multivariate analysis of variance); the purpose still is to compare group means, but now the means of each dependent variable are tested for equivalency across the groups defined by the levels of the independent factor(s). A typical application of a MANOVA in business research might be when a multi-item scale is compared across a few groups; the means of the items in the scale can then be compared simultaneously across groups in a single test, rather than using separate ANOVAs for each item. MANCOVA is a MANOVA that also can consider continuous independent covariates that are analyzed in a fashion similar to regression analysis.

Note that one of the benefits of multivariate statistical methods like MANOVA is that they typically permit the use of a single global statistical test rather than a series of smaller tests, such as running two or more ANOVAs that have different dependent variables, but the same independent variables. Similarly, a single ANOVA can replace a series of t -tests. The simultaneous or global tests are preferred over multiple smaller tests to control the probability of making a Type I error (α). In other words, there is an increased chance of falsely rejecting a null hypothesis (H_o) when using a series of tests versus a single test. The probability of making a Type I error typically is set to $\alpha = 5\%$ (0.05) for each test and, if more than one test is conducted, then α grows to $(1\text{-}(1\text{-}\alpha)^x)$, where x is the number of tests (e.g., for x = three tests, α increases to $1\text{-}(0.95)^3 = 0.143$). Hypothesis testing is not discussed at length here (except for the three fundamental test forms in a subsequent section) because excellent descriptions are available in most or all introductory statistics and research textbooks.

3.2.2. *Logistic regression*

Logistic regression is a special case of regression model whose purpose is to divide or assign the objects of interest (e.g., respondents) into two groups on the basis of one or more independent variables. The decision to classify objects into one group or the other results from the ability of independent variables to differentiate between the two groups, and the best independent variables in a logistic regression will have significantly different means across the groups. Logistic regression often is used with probabilities, because the model limits predictions to a 0–1 range, and a cut point is used to classify each object into one of the two groups. It is because of the focus on means that we categorize logistic regression as a difference method; otherwise, it is used like a typical regression model with one or more independent variables and a single two-category dependent variable.

An example of a logistic regression application could be determining which consumers will or will not purchase a certain product. The model is developed by already knowing into which groups a sample of objects falls (e.g., consumers who did or did not purchase a product), and then determining which independent variables can be used to support the classification. A model typically will not classify every object perfectly, and the model's value and external validity can be assessed by determining how accurately it classifies members of a holdout sample (a sample that was not used during model development).

The method is very similar to discriminant analysis, but logistic regression can classify objects into only two groups, whereas discriminant analysis can be used with two or more groups (although a multi-group version of logistic regression exists, and is known as multinomial logistic regression). However, when faced with the choice between a two-group discriminant analysis or a logistic regression, the latter has the advantage of less restrictive assumptions. Examples of applications include Adusei and Appiah (2011) and Uchenna, Akinola, and Motunrayo (2015).

3.2.3. *Discriminant analysis*

The purpose of discriminant analysis is to find the independent variables that best differentiate between two or more groups. Group membership is the dependent variable, and significant independent variables can accurately discriminate between groups and have significantly different means across groups. Independent variables that do not have significantly different values across groups cannot discriminate between them.

In a sense, discriminant analysis is the opposite of ANOVA, because discriminant analysis has a categorical dependent variable and continuous independent variables; whereas ANOVA has a continuous dependent variable and categorical independent variables. In both cases, significant differences in means are evaluated.

As with logistic regression, the discriminant analysis model is developed with a sample whose group membership already is known, and then a separate holdout sample is used to determine the accuracy of the model in terms of its ability to correctly classify members of the holdout sample into the appropriate groups. Discriminant analysis applications are provided by Eljelly and Mansour (2001) and Gillison, Northington, and Beatty (2014).

3.2.4. *Conjoint analysis*

The notion that products have different characteristics or attributes is widely accepted in marketing. However, the question of which attributes – and how much of them – are most important to customers often arises, because it helps product designers and marketers to create products that stand out from their competitors. Most products have trade-offs between price and quality, or the amount of other desired attributes. For example, cars have attributes such as engine power, gas mileage, steering and handling, interior design and comfort, and electronic features like stereo and navigation. How do researchers determine which attributes and at what level are important to consumers, and what characteristics to emphasize in marketing?

Conjoint analysis is a multivariate statistical method that can be used to compare different characteristics or attributes. In an application such as the car example, respondents view a series of attribute combinations or profiles (e.g., low, medium, or high levels of attributes A, B, and C), and provide rating or preference data so that the value of each attribute can be estimated. The important attributes have higher average ratings or preference figures than less important attributes. Conjoint analysis can be used anytime a researcher wants to compare attributes, whether it be a marketing application or another such as job characteristics. Examples of published applications include Kelley, Hyde, and Bruwer (2015) and Okechuku and Onyemah (2000).

Conjoint effectively is an extension of the MANOVA family because it compares means to determine attribute importance. However, a useful difference between the two methods is that conjoint analysis can not only compare group means, but it also permits estimation of individual-level coefficients for each attribute (part-worths) by obtaining multiple responses from each respondent. In other words, each respondent has their own mean response, and further analysis can be done to explore groups of people with similar responses.

Uso do Solo = atributo peso 5; Clima = 4; Relevo = 3; Unidade de Conservação = 2; Geomorfologia e Fitofissionomia =1
Será? Fazer análise antes de dar pesos para confirmar essa hipótese?

3.3. *The distance (similarity) methods*

The third category of multivariate statistical methods considers the distance between objects or, conversely, their similarity to one another. Similarity can be measured in three ways, including distance, correlations, and association measures, although the distance measures are by far the most frequently used because the correlation and association measures (that effectively are categorical correlations) assess similarities in response profiles rather than similarities in response magnitudes (Hair et al., 2010). The distance methods are represented by cluster analysis, multidimensional scaling, and correspondence analysis, which are seldom applied in the *Journal of African Business* (Table 2).

3.3.1. *Cluster analysis*

This statistical method groups objects (typically respondents) into clusters on the basis of their similarity to one another using variables selected by the researcher. An example might be grouping employees on the basis of their job performance as measured by five job characteristics. One group might have low scores on the first three characteristics, but a high score on the fourth and fifth characteristics. Another group might have high scores on all characteristics, and so on. The number of groups can range from just one to the total number of objects being analyzed (such as the entire sample (N) of employees in the example), but typically is determined by a statistical rule or by the researcher – using theory – prior to the analysis. Clustering methods can proceed using a top down or a bottom up approach. In other words, they can begin with one group, then divide the single group by identifying the two most different groups, and so on, to create N groups or as many groups as needed; or they can begin with N groups and join them, one object at a time, until there is just one group.

After clusters are created, then additional analysis typically is performed to justify the clustering exercise. For example, the clusters could be evaluated to determine whether they are significantly different on the variables that created the clusters, although such an analysis is part of determining how many clusters should be retained. It can be more interesting to determine whether clusters are significantly different from one another on other variables that were not used to create the clusters, particularly if theory is used to create the clusters. In this case, the researcher identifies the expected variable scores and the number of clusters prior to confirmatory analysis; this is known as k -means cluster analysis and Finn, McQuitty, and Rigby (1994) and Winston and Dadzie (2007) provide examples of its use. Marketing makes considerable use of clusters with the notion of segmentation (dividing consumers into groups), which is described as valuable when customer segments (clusters) are different from one another, easy to identify, of meaningful size, and can be used to direct business or marketing actions (e.g., Kotler & Keller, 2016).

3.3.2. *Multidimensional scaling and correspondence analysis*

Both multidimensional scaling (MDS) and correspondence analysis (CA) are used to illustrate the relative locations of comparable objects (e.g., competing products such as cars) and to infer the characteristics or dimensions that people use when comparing the alternatives. However, the primary difference between MDS and CA is the type of data required to estimate the models. Where MDS uses preference or similarity data, CA uses count data for categorical variables (such as brands and product attributes) that can form a cross-table (also known as a contingency table) to provide χ^2 (chi-square) values. Fassin, Rossem, and Buelens (2011) and Gupta and Hanges (2004) are examples of research articles using multidimensional scaling, and Opoku (2009) and Opoku, Pitt, and Abratt (2007) are examples that apply correspondence analysis.

MDS is popular in marketing, for example, where studies ask consumers to rate competing products in terms of their relative preference or similarity. These data can be used to determine the relative locations of products on what is termed a perceptual map, which provides a visual presentation of the competing products. Perceptual maps place the products in n-dimensional space relative to one another (only two dimensions are typically used in illustrations). These positions are then interpreted to infer the bases or dimensions that respondents use to compare the alternatives, and to show, e.g., the degree of competition among the alternative products (closer products are assumed more competitive with one another).

An advantage of MDS is that researchers do not have to specify the characteristics on which alternatives are compared, and respondents create their own bases for comparisons. It also is possible to determine the relative importance of these dimensions to respondents. The mapping and subsequent interpretation can be either group- or individual-level.

For example, if automobile brands are placed on a perceptual map as a result of MDS analysis using similarity data, then it is possible to infer what characteristics or dimensions are used by consumers to compare the brands. If the first dimension has BMW, Mercedes, and Lexus close together while Kia and Chevrolet are far from these three (but close together), then the relative brand positions could be interpreted as a luxury or price dimension. On a second dimension, BMW and Audi might be close together, with Mercedes and Ford further down, and Kia and Lexus still further away. This dimension might be interpreted as sportiness. Knowledge of the relative positions that products have on various dimensions is needed to interpret how respondents compare the alternatives, although researchers must be careful to be objective and not let their own preferences affect the interpretation.

CA also maps the relative position of objects, but uses cross-table data that have two variables, such as products and attributes or age and job performance. Another difference when compared to MDS is that both variables are positioned on the perceptual map, so it is possible to visually determine how one variable is perceived in terms of its relative strengths and weaknesses on the other variable. An example of CA data could use age and restaurant categories to determine if there is a relationship between the two variables. Another example could use products and attributes, where respondents check the attributes they use to evaluate the products. In these cases, the data are counts of the frequencies that one variable category occurs for each category of the second variable.

3.3.3. *Some additional methods*

New ways to analyze data are being developed on an ongoing basis, as computing power increases and the need for different or better tests arises. The relatively newer data analysis methods typically are extensions of existing methods and offer greater model flexibility or have fewer data assumptions (e.g., do not assume that variables are normally distributed). A few important examples are described briefly below.

Generalized linear models (GLMs) are generalizations of regression models that allow dependent variables to have nonnormal distributions (en.wikipedia.org/wiki/Generalized_linear_model), whereas traditional regression models assume that variables are normally distributed. More specifically, GLM was developed to handle situations when the range of the dependent variable is restricted, such as binary or count data, or when the dependent variable's variance is related to its mean (e.g., statmath.wu.ac.at/courses/heather_turner/glmCourse_001.pdf). Logistic regression is a good example of a GLM, but this statistical method covers a number of specialized regression and ANOVA models.

Another newer method is multilevel modeling, which typically is used with nested data and allows for model parameters to vary with different levels of the independent variables (en.wikipedia.org/wiki/Multilevel_model). In other words, multilevel modeling allows the relationships between the dependent and independent variables to change across groups (MANOVA assumes these relationships are constant across groups). An example of nested data is student exam scores as the dependent variable, where students are nested in classrooms as a first independent variable, which are nested in schools as a second independent variable. In this way the effects of classrooms and schools on student exam scores can be evaluated.

A third method that may be gaining in popularity is survival analysis, which estimates the amount of time until an event happens (en.wikipedia.org/wiki/Survival_analysis). Survival analysis estimates the expected duration of time until one or more events happen, such as death in biological organisms or failure in mechanical systems. The method is called reliability theory or reliability analysis in engineering, duration analysis or duration modeling in economics, and event history analysis in sociology. Survival analysis attempts to answer questions such as what is the proportion of a population that will survive past a certain time frame? Of those who/that survive, at what rate will they die or fail? Can multiple causes of death or failure be taken into account? How do particular circumstances or characteristics increase or decrease the probability of survival?

4. The statistical tests underlying multivariate statistical methods

The output from statistical packages can be lengthy, and it sometimes is difficult to know what portion of the output is important. However, it may be comforting to know that all the multivariate statistical methods use one or more of just three types of statistical tests. In other words, there are three basic statistical tests that underlie all statistical analyses (and the three are related mathematically). The three basic statistical tests are the χ^2 (chi-square) test, the t – test, and the F – test. Each of these tests has a different purpose; and when the purpose of each test is understood, then it may be easier to understand the purposes and the output of the various multivariate methods.

The χ^2 test compares two matrices that are the same size. The matrices can be single dimension matrices (arrays), or two dimensional such as when correlation matrices are compared in SEMs, or even multidimensional (e.g., a two dimensional matrix at different points in time). These matrices typically are called the observed matrix and the expected matrix, and the χ^2 test evaluates the differences between them, one element at a time. If the summed differences are large, then the χ^2 test is significant. If the differences between the two matrices are small, then the χ^2 test is not significant. An example of a situation where a χ^2 test is used is a comparison of two categorical distributions, perhaps to determine if two groups have the same distribution of Likert-scale responses. A statistical method that uses the χ^2 test is structural equation modeling, where it provides the basis for goodness-of-fit tests.

The t -test typically is used to evaluate the differences in means across two groups. A t -test also can be used to compare a group mean with a known figure (e.g., zero), to evaluate differences in proportions or percentages across two groups, or to compare a group with a known proportion or percentage. For example, SEM, regression analysis, and various other statistical methods use t -tests to evaluate the significance of individual coefficients. In these cases, the coefficient is compared to a known figure (zero).

F-tests effectively are simultaneous t -tests and also are used to compare means (although the F-test is not always described this way). F -tests are most strongly associated with ANOVAs. If there are more than two groups in a study and the researcher wants to compare the means of all the groups simultaneously (i.e., are the overall differences significantly different), then the F – test should be used. Because using a series of t – tests inflates the probability of a Type I error (i.e., the probability of rejecting H_o when it is "correct" is greater than 5%), a single F -test is preferred. This typically is true of all statistical testing, i.e., a single analysis and test is preferred over a series of analyses and tests (if only to control the probability of making a Type I error (α), as explained in the description of ANOVA). F -tests are found, for example, in the ANOVA family and regression analysis where it is used to test the overall significance of the model (or, effectively, whether a group of coefficients is significantly different from zero).

To summarize, the χ^2 test is used to compare two matrices, the t -test to compare two means (or proportions/percentages as a special case), and the F -test to compare two or more means. It can be difficult to know what matrices or means are being compared. However, if the three fundamental statistical tests are understood, then hypothesis testing and the multivariate statistics should be easier to use and understand. For each of these three tests there are tables to determine whether the calculated test statistic (t, F, or χ^2) is significant. The results obtained from the relevant tests are compared in tables that are available in most statistics texts or online to evaluate significance (e.g., http://www.statsoft.com/textbook/distribution-tables/). Statistical packages typically compute test statistics and associated probabilities for most statistical methods, but some knowledge is required to use the information properly.

5. Multivariate statistical software packages

Many multivariate statistical packages are available, although convenient access, familiarity, price, and features may determine which package works best for an individual researcher. Some packages can conduct most or all the forms of analysis described here and have enough features that they work for all but very specialized analysis. These packages include examples such as SPSS, SAS, JMP, R, Statistica, Matlab, Minitab, PSPP, and Stata. Even Microsoft's Excel can perform many statistical analyses, and some software packages are based on and extend Excel (e.g., XLStat). Specialized software like LISREL, Amos, and EQS exists for analyzing structural equation models (see McQuitty and Wolf (2013) for a further discussion of SEM software), and specialized software can be found for some of the other statistical methods for complex applications.

The various software packages vary in price considerably. Some of the software packages are available for free, such as PSPP (https://www.gnu.org/software/pspp/) or R (https://www.r-project.org/). For university professors and students, SAS has a free version (sas.com/university-edition) that appears able to handle all the methods described in this commentary and more. Other packages can be somewhat expensive, such as IBM's SPSS (http://www.ibm.com/analytics/us/en/technology/spss/), particularly when additional features add to the price. However, any expense is justifiable if the software package does analysis easily and well.

In the past, researchers' choices of statistical analysis software may have depended on availability and their statistical background and ability to program (e.g., Gallaher, Ting, & Palmer, 2008). Most packages today use menus so their use is relatively straightforward and no programming is necessary. SPSS, for example, uses menus and tends to have many options for statistical analysis, so the emphasis should be on understanding the effects of all the options, such as estimation method, ways of handling missing data, etc. The output from multivariate analysis can be lengthy, so understanding and correctly interpreting the output is critical.

6. Concluding comments

Attempting to introduce and summarize the purpose of multivariate statistical methods in a brief commentary is, in many ways, a reckless undertaking due to the complicated nature of the topic . However, formal statistics education tends to focus on specific knowledge like mathematical relationships and probability, rather than the application of the statistical methods. Mathematics and probability clearly underlie all statistical analysis, but it is unnecessary to know these topics well to use modern statistical analysis software packages, so the bigger issue is the choice of appropriate methods.

Therefore, this introduction is intended for readers who have some statistical background and are aware that a variety of multivariate methods exist, but do not have extensive training in these methods nor understand how to use them. For these people, an explanation of the purpose of each multivariate method may be enough to help them select an appropriate method to analyze their research problem. It will be necessary to read more about the chosen statistical method to obtain the specific information needed to use the method properly. The statistical method must be consistent with the research question and the data required to properly evaluate the research question, so these all should be considered as part of the research design prior to data collection.

Be aware that, to some degree, each business discipline will have a preferred way of describing the use of statistical methods, so researchers are encouraged to examine published articles in their areas to learn the expected approach, and particularly articles from the journals to which they submit manuscripts. There may be published articles that describe the best way to write about specific multivariate methods; a good example is SEM, with articles that explain how to report a SEM application (e.g., Hoyle & Panter, 1995) or studies of how well SEM has been applied (e.g., Baumgartner & Homburg, 1996; Hulland, Chow, & Lam, 1996).

References

Adusei, M., & Appiah, S. (2011). Determinants of group lending in the credit union industry in Ghana. *Journal of African Business*, *12*(2), 238–251. doi:10.1080/15228916.2011.588914

Albinsson, P. A., Yasanthi, P. B., & Sautter, P. T. (2016). DART scale development: Diagnosing a firm's readiness for strategic value co-creation. *Journal of Marketing Theory and Practice*, *24*(1), 42–58. doi:10.1080/10696679.2016.1089763

Batista-Foguet, J. M., Coenders, G., & Ferragud, A. (2001). Using structural equation models to evaluate the magnitude of measurement error in blood pressure. *Statistics in Medicine*, *20*, 2351–2368. doi:10.1002/sim.836

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing*, *13*(2), 139–161. doi:10.1016/0167-8116(95)00038-0

Biboum, A. D., & Sigué, S. P. (2014). Conflict in supplier–retailer relationships in the brewery industry in Cameroon. *Journal of African Business*, *15*(2), 75–84. doi:10.1080/15228916.2014.925361

Blankson, C., Mbah, C. H., & Owusu-Frempong, L. M. (2009). The development of a scale of retail consumers' selection of retail banks in Ghana. *Journal of African Business*, *10*(2), 182–202. doi:10.1080/15228910903187742
Cattell, R. B. (1952). The three basic factor-analytic research designs—Their interrelations and derivatives. *Psychological Bulletin*, *49*, 499–520. doi:10.1037/h0054245
Chen, H., Chen, P., & Okumus, F. (2013). The relationship between travel constraints and destination image: A case study of Brunei. *Tourism Management*, *35*, 198–208. doi:10.1016/j.tourman.2012.07.004
Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, *16*(1), 64–73. doi:10.2307/3150876
Dadzie, C. A., Winston, E. M., & Dadzie, K. Q. (2012). Organizational culture, competitive strategy and performance in Ghana. *Journal of African Business*, *13*(3), 172–182. doi:10.1080/15228916.2012.727737
Eljelly, A. M. A., & Mansour, I. H. F. (2001). Predicting private companies failure in the Sudan. *Journal of African Business*, *2*(2), 23–43. doi:10.1300/J156v02n02_03
Fassin, Y., Rossem, A., & Buelens, M. (2011). Small-business owner-managers' perceptions of business ethics and CSR-related concepts. *Journal of Business Ethics*, *98*(3), 425–453. doi:10.1007/s10551-010-0586-y
Finn, A., McQuitty, S., & Rigby, J. (1994). Area resident acceptance and use of mega-multi-malls: West Edmonton Mall evidence. *International Journal of Research in Marketing*, *11*(1), 127–144. doi:10.1016/0167-8116(94)90024-8
Gagoitseope, P. K., & Pansiri, J. (2012). Evaluation of critical success factors for developing small and medium-sized enterprises in Botswana. *Journal of African Business*, *13*(1), 51–61. doi:10.1080/15228916.2012.657955
Gallaher, D., Ting, L., & Palmer, A. (2008). A journey into the unknown: Taking the fear out of structural equation modeling with AMOS for the first-time user. *The Marketing Review*, *8*(3), 255–327. doi:10.1362/146934708X337672
Gillison, S. T., Northington, W. M., & Beatty, S. E. (2014). Understanding customer bargaining in retail stores: A customer perspective. *Journal of Marketing Theory and Practice*, *22*(2), 151–168. doi:10.2753/MTP1069-6679220210
Grace, J. B. (2006). *Structural equation modeling and natural systems*. New York: Cambridge University Press.
Gupta, V., & Hanges, P. J. (2004). Regional and climate clustering of societal cultures. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.), *Culture, leadership, and organizations: The GLOBE study of 62 societies* (pp. 178–218). Thousand Oaks, CA: Sage Publications.
Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
Hair, J. F., Jr., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H.Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage Publications.
Hulland, J., Chow, Y. H., & Lam, S. (1996). Use of causal models in marketing research: A review. *International Journal of Research in Marketing*, *13*(2), 181–197. doi:10.1016/0167-8116(96)00002-X
Kamukama, N. (2013). Intellectual capital: Firms' hidden source of service quality in the microfinance industry in Uganda. *Journal of African Business*, *14*(3), 150–161. doi:10.1080/15228916.2013.844012
Kelley, K., Hyde, J., & Bruwer, J. (2015). U.S. wine consumer preferences for bottle characteristics, back label extrinsic cues and wine composition: A conjoint analysis. *Asia Pacific Journal of Marketing and Logistics*, *27*(4), 516–534. doi:10.1108/APJML-09-2014-0140
Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Boston: Pearson.

Makanyeza, C., & du Toit, F. (2016). Measuring consumer ethnocentrism: An assessment of reliability, validity and dimensionality of the CETSCALE in a developing market. *Journal of African Business*, *17*(2), 188–208. doi:10.1080/15228916.2016.1138270

McQuitty, S., & Wolf, M. (2013). Structural equation modeling: A practical introduction. *Journal of African Business*, *14*(1), 58–69. doi:10.1080/15228916.2013.765325

Mori, N. (2014). Directors' diversity and board performance: Evidence from East African microfinance institutions. *Journal of African Business*, *15*(2), 100–113. doi:10.1080/15228916.2014.920654

Musila, J. W. (2013). Does democracy have a different impact on corruption in Africa?*Journal of African Business*, *14*(3), 162–170. doi:10.1080/15228916.2013.844020

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Okechuku, C., & Onyemah, V. (2000). Ethnic differences in Nigerian consumer attitudes toward foreign and domestic products. *Journal of African Business*, *1*(2), 7–35. doi:10.1300/J156v01n02_02

Opoku, R. A. (2009). Mapping destination personality in cyberspace: An evaluation of country web sites using correspondence analysis. *Journal of Internet Commerce*, *8*(1/2), 70–87. doi:10.1080/15332860903182438

Opoku, R. A., Pitt, L. F., & Abratt, R. (2007). Positioning in cyberspace: Evaluating bestselling authors' online communicated brand personalities using computer-aided content analysis. *South African Journal of Business Management*, *38*(4), 21–32.

Osinubi, T. S., & Amaghionyeodiwe, L. A. (2003). Stock market development and long-run growth in Nigeria. *Journal of African Business*, *4*(3), 103–129. doi:10.1300/J156v04n03_06

Osuagwu, L. (2002). The marketing research practices of Nigerian companies. *Journal of African Business*, *3*(1), 81–107. doi:10.1300/J156v03n01_05

Parasuraman, A., Ziethaml, V., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *62*(1), 12–40.

Pugesek, B. H., Tomer, A., & von Eye, A. (Eds.). (2003). *Structural equation modeling applications in ecological and evolutionary biology*. Cambridge: Cambridge University Press.

Rothmann, S., & Rothmann, S., Jr. (2010). Factors associated with employee engagement in South Africa. *South African Journal of Industrial Psychology*, *36*(2), 27–38.

Sigué, S. P., & Bonsu, S. K. (2012). Influence strategies and channel member satisfaction in Cameroon. *Journal of African Business*, *13*(3), 200–208. doi:10.1080/15228916.2012.727747

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Uchenna, E., Akinola, O., & Motunrayo, O. (2015). Workers' deviant behavior in the business outfit of a Nigerian university: Situating policies with empirics. *Journal of African Business*, *16*(3), 339–357. doi:10.1080/15228916.2015.1090829

Winston, E., Dadzie, C. A., & Dadzie, K. Q. (2009). How managers handle conflict in supply chain collaborative relationships in Ghana. *Journal of African Business*, *10*(2), 203–217. doi:10.1080/15228910903187833

Winston, E., & Dadzie, K. Q. (2007). Organizational culture in Cote d'Ivoire: Analysis of cultural types and firm performance. *Journal of African Business*, *8*(1), 99–112. doi:10.1300/J156v08n01_06

Zoogah, D. B. (2008). African business research: A review of studies published in the Journal of African Business and a framework for enhancing future studies. *Journal of African Business*, *9*(1), 219–255. doi:10.1080/15228910802053037