

Multivariate Data Analyses:
Multiple Regression Applied in Educational Research

by

Erlinda P. Villamoran

For additional information, please contact:

Author’s name : Erlinda P. Villamoran
Designation : University Professor
Affiliation : Bulacan State University
Address : City of Malolos
Tel. no. : (0917) 896-2542
E-mail : erlinda.villamoran@yahoo.com
Multivariate Data Analyses:  
Multiple Regression Applied in Educational Research  
by  
Erlinda P. Villamoran  

ABSTRACT  
Multivariate analysis has its roots in univariate and bivariate statistics, the extension to the multivariate domain introduces additional concepts and issues that have peculiar relevance. These concepts range from the need for a conceptual understanding of the basic building block of multivariate analysis - the variate to specific issues dealing with the types of measurement scales used and the statistical issues of significance testing and confidence levels.  

This paper will focus on three main concepts. First the objectives of multiple regression which will consider three primary issues: (1) the appropriateness of the research problem, (2) specification of a statistical relationship and, (3) selection of the independent and dependent variables.  

The ever-widening applications of multiple regression fall into two broad class of research problems: predictions and explanations. These research problems are met mutually exclusive, and an application of multiple regression analysis can address either or both types of research problem.  

One fundamental purpose of multiple regressions is the prediction of the dependent variable with a set of independent variables. It provides an objective means of assessing the predictive power of a set of independent variables. Research agenda will be presented in the paper with the corresponding analysis of statistical techniques to be applied, specifically, the multiple regression analysis.  

The paper will also present the research design of a multiple regression analysis; the assumptions underlying the meaning of the relationships between dependent and independent variables. There will be an illustration of a Regression Analysis focus on 6 stages: (1) objectives of the multiple regression, (2) research design of the multiple regression analysis, (3) assumptions of the multiple regression analysis, (4) estimating the regression model and assessing overall model fit, (5) interpreting the variate and, (6) validating the results. Specific samples will be used to represent the illustrations.  

Introduction  
Technology advances are very evident globally. The computer technology for example has made possible extraordinary advances in the analysis of psychological, sociological, and other types of behavioral data. This impact is most evident in the relative ease with which computers can analyze large quantities of complex data. Several statistical programs are available applicable to some specific disciplines. Educational researchers can easily manipulate data, test hypothesis and make statistical inferences. This application of statistical techniques is known as multivariate analysis. They are widely used in industry, agriculture, government, and university – related research centers, e.g. educational researches.
Although two variable relationships can be used in research, multivariate analysis is a more powerful tool. Some noted statisticians said that unless a problem is tested as a multivariate problem, it is tested superficially. Therefore, we in the education sectors should learn the basics and application of this statistical technique to educational researches.

**Theoretical framework**

Many application of inferential statistics are much more complex than the methods used in the investigation of relationships between two variables. If we investigate the relationships among a group of variables, we have to create a model for some variables that can be used to predict its value in the future. The process of finding a mathematical model (an equation) that best fits the data is part of a statistical technique known as regression analysis. Regression analyses are sets of statistical techniques that allow one to assess the relationship between one dependent variable (DV) and several independent variables (IVs). Regression is often used when the intent of the analysis is prediction, and the term correlation is used when the intent is simply to assess the possible linear relationship between the DV and the IVs.

Most practical applications of regression analysis utilize models that are more complex than the first – order (straight-line) model which is shown in the box

<table>
<thead>
<tr>
<th>A First-Order Straight Line Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y = B_0 + B_1 \cdot x + e )</td>
</tr>
<tr>
<td>( y = ) dependent variable (criterion)</td>
</tr>
<tr>
<td>( x = ) Independent Variable (predictor)</td>
</tr>
<tr>
<td>( E(y) = B_0 + B_1 x = ) deterministic component</td>
</tr>
<tr>
<td>( e = ) (epsilon) = random error component</td>
</tr>
<tr>
<td>( B_0 = (beta \ zero) = y ) – intercept of the line</td>
</tr>
<tr>
<td>( B_1 = (beta \ one) = Slope \ of \ the \ line, )</td>
</tr>
<tr>
<td>( ) (amount in increase (or decrease) in the mean of ( y ) for every ( 1 ) unit increase in ( x ).)</td>
</tr>
</tbody>
</table>

Empirical Issues in the Decision Process for Multivariate Analyses

Stage 1. Define The Research Problem, Objectives and Multivariate to be Used:

1.1 Is the problem suitable for multivariate analysis?
1.2 Which multivariate technique is best suited to the research problem?

Define the research problem and analyze objectives in conceptual terms before specifying any variables or measures. Define the concepts and
identify the fundamental relationships to be investigated. A conceptual model need not be complex and detailed but a sample representation of the relationships to be studied. If a dependence relationship is proposed as the research objective, the researcher needs to specify the dependent and independent concepts. For an application of an interdependence technique, the dimension of structure or similarity should be specified. In both the dependence and interdependence situations, the researcher first identifies the ideas or topics of interest rather than focusing on the specific measures to be used.

Stage 2. Develop and Analysis Plan

2.1 How does sample size affect your results?

2.2 Are the variables of the correct measurement type? If not, can they be transformed?

Develop a specific analysis plan that addresses the set of issues particular to its purpose and design. The issues range from the general considerations of minimum or desired sample sizes, to allowable or required types of variables.

Stage 3. Evaluate the Assumptions of the Multivariate Technique

3.1 Have the underlying assumptions been tested empirically?

With the data collected, evaluate the underlying assumptions. For the techniques based on statistical inference, the assumptions of multivariate normality, linearity, independence of the error terms, and equality of variances in a dependence relationship must all be met.

Stage 4. Estimate the Multivariate Model and Assess Overall Model Fit

4.1 What is the statistical power of the multivariate technique?

4.2 How do you interpret the errors of prediction or explanation?

With the assumptions satisfied, the analysis proceeds for the actual estimation of the multivariate model and an assessment of the overall model fit.

Stage 5. Interpret the Variate

5.1 Are the results evaluated with some measure of statistical significance?

5.2 What results came from the variate versus evaluating individual variable(s)?
Interpreting the variate (s) reveals the nature of multivariate relationship.

Stage 6. Validate the Model

6.1 How do you compare and evaluate the differing results obtained in your validation efforts?

Validate the model by demonstrating the generalizeability of the results to the total population.

Probabilistic model that includes terms involving $x^2$, $x^3$ (or higher – order terms), or more than one independent variable are called multiple regression models or linear statistical models. The goal of regression is to arrive at the set of $B$ values called regression coefficients for the IVs that bring the Y values predicted from the equation as close as possible to the Y values obtained by measurement. The regression coefficients that are computed accomplish two intuitively appealing and highly desirable goals. They minimize (the sum of the squared) deviations between predicted and obtained Y values for the date set. The Pearson product-amount correlation coefficient between the obtained and predicted Y values was derived from a regression analysis, termed as multiple correlation coefficient ($R$).

Regression techniques consist of standard multiple regression, sequential (hierarchial) regression and statistical (stepwise) regression. This presentation will focus only on the standard multiple regression. The general forms of this model is shown in the box

<table>
<thead>
<tr>
<th>The General Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y = B_0 + B_1x_1 + B_2x_2 + ... B_kx_k + \theta$</td>
</tr>
<tr>
<td>Where: $y$ is the dependent variable</td>
</tr>
<tr>
<td>and $x_1, x_2 ... x_k$ are the independent variables.</td>
</tr>
</tbody>
</table>

$$E(y) = B_0 + B_1x_1 + ... + B_kx_p$$ is the deterministic portion of the model

$B_0 = y$ intercept

$\theta = \text{random errors’ component}$

$B_1$ determines the contribution of the Independent variable $X_i$. 


1. In choosing the IVs, regression will be best when each IV is strongly correlated with the DV but uncorrelated with other IVs. A general goal of regression, then, is to identify the fewest IVs necessary to predict a DV were each IV predicts a substantial and independent segment of the variability in the DV. Regression analyses can be used with either continuous or dichotomous IVs while the DV should be continuous. A variable that is initially discrete can be used if it is first converted into a set of dichotomous variables (numbering one fewer than the number of discrete categories) by dummy variable coding with 1’s and 0’s, e.g. assessing religious affiliation in which 1 stands for Protestant, 2 for Catholic C, 3 for INK. The variable may be converted into a set of three new variables, (Protestant/non-Protestant, Catholic/non-Catholic, etc).

2. The required sample size in running a regression analysis depends on a number of issues, including the desired power, alpha level, number of predictors and expected effect sizes. Green (1991) provides a simple rule of thumb, for testing the multiple correlation, $N = 50 + 8m$ (where $m$ is the number of IVs and $N = 104 + m$ for testing individual predictors).

3. Calculation of regression coefficients requires inversion of the matrix of correlations among the IVs. They are unstable if the IV’s are multicollinear. A multi-collinearity among the IVs also signaled very large standard errors for regression coefficients. Berry (1993) reports that when $r$ is 0.9, standard errors of the regression coefficients are doubled; when multicollinearity is present, more of the regression coefficients may be significant because of the large size of standard errors. In this case, some of the highly correlated IVs may be deleted using logical rather than statistical grounds by considering issues such as the reliability of the variables or the cost of measuring the variables.

On the other hand, if multicollinearity is detected but you want to maintain your set of IVs, ridge regression might be considered. These regressions stabilize estimates of regression coefficients by inflating the variance that is analyzed.

Analysis of variances is a part of the general linear model, as is regressions.
Some applications of the multiple regression analysis is included in this presentation.
References


Statement of the Problem

The general problem of the study is: How do the emotional quotient and work attitudes affect the leadership skills of administrators, and consequently the job satisfaction and job productivity of the faculty of the state agricultural institutions in Region III during the school year 2002-2003?

Specifically, this study sought answers to the following questions:

1. How may the profile of the respondents be described in terms of:
   1.1. administrators profile;
      1.1.1. age;
      1.1.2. gender;
      1.1.3. civil status;
      1.1.4. highest educational attainment;
      1.1.5. length of service as administrator;
   1.2 faculty profile;
      1.2.1 age;
      1.2.2 gender;
      1.2.3 civil status;
      1.2.4 highest educational attainment;
      1.2.5 length of service as teacher; and
      1.2.6 academic rank?

2. What is the level of the emotional quotient and work attitude of the administrator?

3. What is the level of leadership skills of the administrator as perceived by the faculty members with regards to:
   3.1 technical;
   3.2 human; and
   3.3 conceptual?

4. Are there significant effects of the emotional quotient and work attitudes of the administrator on their leadership skills?

5. How does the profile of the administrators influence the effects of emotional quotient and work attitudes on their leadership skills?

6. What is the level of job satisfaction and job productivity of the faculty members viz-a-viz;
   6.1 job satisfaction;
      6.1.1 physiological needs;
      6.1.2 social needs; and
6.1.3 intellectual needs;
6.2 job productivity;
   6.2.1 institutional competencies;
   6.2.2 personal characteristics;
   6.2.3 human relations;
   6.2.4 classroom management;
   6.2.5 punctuality and accuracy;
   6.2.6 teaching procedures;
   6.2.7 professional attitudes; and
   6.2.8 interaction dynamics?

7. Are there significant effects of the leadership skills of the administrators on the job satisfaction and job productivity of the faculty members?

8. How does the profile of the faculty members influence the effects of leadership skills of the administrators on their job satisfaction and job productivity?
Figure 2. The Conceptual Model of the Study
Hypotheses of the Study

The following null hypothesis were tested in this study:
1. The administrators’ emotional quotient and work attitude do not significantly affect their leadership skills, measured in terms of technical, human, and conceptual.
2. The profile of the administrators does not significantly influence the effects of emotional quotient and work attitude on their leadership skills.
3. The leadership skills of the administrator do not significantly affect the job satisfaction of the teachers.
4. The leadership skills of the administrators do not significantly affect the job productivity of the teachers.
5. The profile of the teachers do not significantly influence the effects of leadership skills of the administrators on the job satisfaction of the teachers.
6. The profile of the teachers do not significantly influence the effects of leadership skills of the administrator on the job productivity of the teachers.

Table 1
Regression Analysis of the Technical Leadership Skills of Administrators On their Emotional Quotient and Work Attitudes

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>T</th>
<th>Sig t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Quotient</td>
<td>.334156</td>
<td>3.894</td>
<td>.0001**</td>
</tr>
<tr>
<td>Work Attitudes</td>
<td>.173658</td>
<td>2.023</td>
<td>.0448*</td>
</tr>
</tbody>
</table>

F = 18.91607, Sig F = .0000**, R² = .2

* p < 0.05; significant
** p < 0.01; highly significant

Table 2
Regression Analysis of the Technical Leadership skills of Administrators on their Emotional Quotient, Work Attitudes Controlling for Profile of Administrators

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>T</th>
<th>Sig t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Quotient</td>
<td>.303472</td>
<td>.755</td>
<td>.4696</td>
</tr>
<tr>
<td>Work Attitudes</td>
<td>-.315882</td>
<td>-.733</td>
<td>.4822</td>
</tr>
<tr>
<td>Age</td>
<td>.353737</td>
<td>1.125</td>
<td>.2899</td>
</tr>
<tr>
<td>Sex</td>
<td>-.205892</td>
<td>-.525</td>
<td>.6121</td>
</tr>
<tr>
<td>Civil Status</td>
<td>-.106850</td>
<td>-.225</td>
<td>.8268</td>
</tr>
<tr>
<td>Educational Attainment</td>
<td>.116207</td>
<td>.278</td>
<td>.7874</td>
</tr>
<tr>
<td>Length of Service</td>
<td>-.705724</td>
<td>-1.675</td>
<td>.1283</td>
</tr>
</tbody>
</table>
### Table 3
Regression Analysis of the Human Leadership Skills of Administrators on their Emotional Quotient, Work Attitudes Controlling for Profile of Teachers

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>T</th>
<th>Sig t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotional Quotient</td>
<td>.385693</td>
<td>1.573</td>
<td>.1502</td>
</tr>
<tr>
<td>Work Attitudes</td>
<td>-.153478</td>
<td>-.584</td>
<td>.5736</td>
</tr>
<tr>
<td>Age</td>
<td>.192209</td>
<td>1.002</td>
<td>.3427</td>
</tr>
<tr>
<td>Sex</td>
<td>.033426</td>
<td>.140</td>
<td>.8919</td>
</tr>
<tr>
<td>Civil Status</td>
<td>-.636104</td>
<td>-2.198</td>
<td>.0500*</td>
</tr>
<tr>
<td>Educational Attainment</td>
<td>-.014318</td>
<td>-.056</td>
<td>.9565</td>
</tr>
<tr>
<td>Length of Service as Teacher</td>
<td>-.986261</td>
<td>-3.837</td>
<td>.0040**</td>
</tr>
</tbody>
</table>

F = 5.00586  
Sig. F = .0134*  
R² = .817  
*p < 0.05, significant  
**p < 0.01, highly significant

### Table 4
Regression Analysis of Job Satisfaction of the Faculty Members in terms of Physiological Needs on the Leadership Skills of the Administrators

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>T</th>
<th>Sig t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Skills</td>
<td>.332418</td>
<td>2.210</td>
<td>.0286*</td>
</tr>
<tr>
<td>Human Skills</td>
<td>.427721</td>
<td>2.768</td>
<td>.0064**</td>
</tr>
<tr>
<td>Conceptual Skills</td>
<td>.24824</td>
<td>4.585</td>
<td>.0000**</td>
</tr>
</tbody>
</table>

F = 11.69029  
Sig. f = .0000**  
R² = .192  
*p < 0.05, significant  
**p < 0.01, highly significant

### Table 5
Regression Analysis of the Job Satisfaction of Faculty Members in terms of Physiological Needs on the Leadership Skills of Administrators and Faculty Members' Profile

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>T</th>
<th>Sig t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Skills</td>
<td>.399379</td>
<td>2.220</td>
<td>.0283*</td>
</tr>
<tr>
<td>Human Skills</td>
<td>.423698</td>
<td>2.340</td>
<td>.0210*</td>
</tr>
<tr>
<td>Conceptual Skills</td>
<td>-.398082</td>
<td>-2.310</td>
<td>.0226*</td>
</tr>
<tr>
<td>Age</td>
<td>.100544</td>
<td>.610</td>
<td>.5433</td>
</tr>
<tr>
<td>Sex</td>
<td>.108702</td>
<td>1.270</td>
<td>.2067</td>
</tr>
<tr>
<td>Civil Status</td>
<td>-.072272</td>
<td>-.827</td>
<td>.4101</td>
</tr>
<tr>
<td>Educational Attainment</td>
<td>-.038599</td>
<td>-.304</td>
<td>.7620</td>
</tr>
<tr>
<td>Length of Service as Faculty</td>
<td>-.149120</td>
<td>-.777</td>
<td>.4387</td>
</tr>
</tbody>
</table>
Academic Rank | -.006445 | -.044 | .9651
F = 3.25394 | Sig. f = .0010** | R² = .215
* p < 0.05, significant **p<0.01, highly significant

Fig. 1 Conceptual Model of the Study

Grammar Proficiency
1. Correct Grammatical Form
2. Answering Questions
3. Combining Sentences
4. Word Sequencing

Hypotheses of the Study

Language attitude, motivational intensity and learning and thinking styles have no significant effects on grammar proficiency.