CHAPTER 5

CASE DATA ANALYSIS

CAUSES & EFFECTS FOR PROVINCIAL SPENDING ON HIV/AIDS

5.1 INTRODUCTION

In chapter 3, section 3.3, variable sets were presented that included data pertaining to nine (9) provinces. This chapter 5 encompasses an analysis of that data, with the aim of either confirming or refuting the hypothesis that voters are unable to influence public policy on provincial government spending for HIV/AIDS. As in other research, the expected direction, negative or positive, and effect on the dependent variable for provincial government spending is indicated. Linear relationships of a positive or negative nature are realised upon running a variety of bivariate and multivariate regressions in order to arrive at an optimal model. Correlation coefficients indicate the strength of the linear relationship and coefficients of determination indicate the degree of predictability of the independent variable. Test of hypothesis are conducted to further confirm that there is a linear relationship between two variables in the optimal bivariate model. The chapter concludes by refuting the hypothesis, while commenting briefly on policy implications.

Husted and Kenny's approach (1997:64) to empirical analysis was to first state the expectations of their theory. In other words, they would state a theory (or a hypothesis) and their subsequent data analysis would go on to substantiate or refute their expectations – i.e., their theory. In that manner, this dissertation examines the theory that there is some relationship between voter turnout and public policy making for HIV/AIDS

treatment. Notably, policy making at the provincial sphere of government for HIV/AIDS is understood to be reflected by provincial government spending for HIV/AIDS. The hypothesis stated earlier was that voters do not have the potential to influence public policy decisions when they exercise their franchise to vote. It is not implied that there is not a relationship. Rather, the strength of the relationship between voters and public policy decision-making remains to be determined but nevertheless expected to be minimal. Upon this data analysis, there may very well be a near negative relationship between voter turnout and national and provincial spending for HIV/AIDS. It is not expected that there would be a strong positive relationship between the two variables. In other the words, in keeping with the hypothesis, an increase in government spending should not be expected to be a significant result of voter turnout. There should, perhaps, be some small influence of the variable VOTE_TURN_04 on PROV_SPEND_03. This is measured by examining the bivariate relationship by way of ordinary least squares (OLS) or simple least squares calculated to determine the relationship between the two primary variables – namely, VOTE_TUR_04 the independent variable and PROV_SPE_03 the dependent variable. Additional calculated parameters include the correlation coefficient (R) to evaluate the strength of the linear relationships.

The data analysis should be qualified by noting that increases in government spending will occur in anticipation of and prior to elections – say, the 2004 elections – in order to garner political support of political incumbents. However, once elected there is uncertainty as to the responsiveness of incumbents due to the party list electoral system that promotes first to the incumbent's political party and then to constituents.

Other effects [variables] are measured against national and provincial spending as well. Regression coefficients will give indication of the strength of the relationship between an independent variable and a dependent variable in the case of bivariate regression, and the strength of the relationship between several independent variables on the dependent variables in the case of multivariate regression. Data output relating to tolerance and multicollinearity (Norusis, 1998:467) will be analysed to determine the strength of any linear relationship among independent variables. A value close to zero [0] indicates that a variable is linear to another independent variable. It might then be necessary to exclude the variable from regression model.

Reference is made to Table 3.2 that includes 23 independent variables thought to have some predictive-effect on public policy decision making as reflected in government spending for HIV/AIDS. In the case of bivariate and multivariate relationships, the ultimate goal is to build a model that offers explanatory effects for government spending on HIV/AIDS. Several regression models will then be considered, some variables in the models eliminated, and some models deemed inadequate towards drawing a conclusion. For the bivariate regressions to be calculated using SPSS, the following equation is offered:

$$Y_i = \boldsymbol{b}_0 + (b_1 \cdot X_1) + e_i$$

And for the multivariate regression the following equation is offered:

$$Y = a + (b_1 \cdot X_1) + (b_2 \cdot X_2) + (b_3 \cdot X_3) + \ . \ \ . \ \ . \ (b_n \cdot X_n)$$

117

where:

Y = The Dependent Variable Provincial Government Spending for HIV/AIDS

a /b = The *Y*-Intercept

$b_1, b_2, b_3 \ldots b_n$ = The Partial Slope Indicating the Linear Relationship Between A Specific Independent Variable and the Dependent Variable

$X_1, X_2, X_3 \ldots X_n$ = A Specific Independent Variable Thought To Be An Efficient Predictor

Notably, the equations above were presented and discussed in chapter three. The variables were defined in sections 3.4.2, 3.4.3, 3.4.4, and 3.4.5. Indeed, section 3.4 of that chapter discussed the statistical approach and techniques to be used, thereby finalising this approach performing calculations using SPSS and subsequent data analysis. As a review:

- Descriptive Statistics are calculated for all independent variables. Parameters include minimum-maximum, the mean and standard deviation.

- Linear Regression (Bivariate Relationships): Ordinary least squares (OLS) or simple least squares are calculated to determine the relationship between the two primary variables – namely, VOTE_TUR_04 the independent variable and PROV_SPE_03 the dependent variable. Other bivariate relationships are examined as well. Additional calculated parameters include the correlation coefficient (*R*) to evaluate the strength of the linear relationships.

- Multiple Regression (Multivariate Relationships): Where in OLS the relationship between two variables is examined, in predicting the value of the dependent variable, say, PROV_SPE_03 or any dependent variable for that matter, the dependent variable is regressed on a number of independent variables (multiple linear regression), recognising that in the real world there are many factors impacting and influencing a dependent variable. While searching for the optimal regression model, there is an examination of tolerances to eliminate variables that are being influenced by other independent variables

- Testing the Regression Hypothesis (Test of Hypothesis): A test of hypothesis is run to determine the representative-ness of the regression straight-line. Is the regression [straight] line adequate for predictability? As mentioned earlier, with the sample size being will be small (nine) and with no population to consider the test of hypothesis is construed to represent the predictability of the sample regression.

- Test of Hypothesis for Reliability of Correlation Coefficient: As a further test of the primary bivariate regression *Fisher's Z transformation* (see chapter 3, p.82) is used to test the reliability of the correlation coefficient R. Thus, a rejection of the null hypothesis indicates R is not a significant predictor of the linear relationship between VOTE_TURN_04 the independent variable and PROV_SPEND_03 the dependent variable.

- Conclusion Drawn from Outputs: This chapter [five] concludes by drawing a final conclusion as to the predictability of bivariate and multivariate regression models. The hypothesis that voters do not have the ability to influence government spending for HIV/AIDS will be confirmed or refuted.

Heretofore, each section that follows presents the data output that are the result of statistical techniques described above. Moreover, the data output are discussed, analysed and interpreted to draw a final conclusion relating to the stated hypothesis.

5.2 DESCRIPTIVES

Table 5.1 presents SPSS output summarising descriptive statistics for variables defined in Table 3.2. The mean provincial voter turnout for the 2004 elections was 1.7 million voters with a standard deviation close to 950 thousand voters. Notably, in 1999 the mean for registered voters was 1.9 million and in 2004 the mean for registered voters increased to 2.29 million voters. The average increase in voter registration of nearly 19%, along with average provincial special interest [TAC] activism of 11% may indicate some latent force at work enabling the electorate to influence policy making through the voting franchise. With that in mind, the minimum voter turnout amongst South Africa's nine provinces was nearly 319 thousand voters and the maximum turnout was more than 3.4 million voters. The vast difference between the minimum and the maximum is a reflection of the distribution of the population, with a remote province being sparsely populated and a more central province being densely populated. In 2001, the mean total provincial population (age 15 – 65) was 4.98 million; by 2003 the average provincial population had increased to 5.23 million people, on average growing 6% over 2 years.

Table 5.1

SPSS Output Descriptives

	N	Minimum	Maximum	Mean	Std. Deviation
VOTE_TURN _04	9	318702	3408308	1700349.11	948632.222
WH_RACE_01	9	.02	.41	.1112	.12360
EDUCA_01	9	145344	2055855	871780.56	585930.247
INC_01	9	.06	.30	.1848	.08358
AIDS_PREV_02	9	.04	.18	.1301	.04728
NEED_04	9	250000	2862000	1480333.33	846793.363
LAT_GROUP_04	9	.08	.41	.1896	.09448
SPEC_INT_01	9	.00	.49	.1122	.16071
CHANGE_PROV_GDP_03	9	-.54	1.35	.0496	.59904
PROV_SPEND_03	9	11300000	246500000	76666666.67	75760279.831
REG_VOTERS_99	9	368205	4119164	1934779.11	1135971.570
REG_VOTERS_04	9	433591	4650594	2297214.00	1305563.893
NATL_SPEND_02	9	.23	8.43	3.3344	2.36410
TOT_POP_01	9	822727	9426017	4980197.33	2843162.358
TOT_POP_03	9	1011774	9556833	5233926.11	2908324.203
POP_GROW_03	9	.01	.19	.0623	.05691
PROV_GDP_02	9	.0146	.0487	.030511	.0128755
PROV_GDP_03	9	.0177	.0386	.026711	.0069449
DEM_GOVSERV_02	9	-.0094	.0730	.014833	.0244581
DEM_GOVSERV_03	9	.0007	.0780	.014489	.0243638
CHANGE_ DEM_GOVSERV_03	9	-1.6596	35.0000	5.362837	12.2266389
NNP_RACE_04	9	.0043	.1088	.024978	.0389391
Valid N =	9				

Source: SPSS Output Computed for Descriptive Statistics

On average, provincial expenditures for HIV/AIDS in 2003 amounted to R76.6 million rands, with a standard deviation of R75.7 million rands. The minimum amount spent by a province was R11.3 million rands and the maximum amount spent by a province was R246.4 million rands. Conditional grants made to provinces in 2002 represent national spending efforts to alleviate HIV/AIDS. On average, in 2002 the national government provided R3.3 million rands in grants targeted for HIV/AIDS. The standard deviation

was R2.4 million rands, with a minimum grant of R .23 million rands and a maximum grant of R8.4 million rands. Notably, by 2002 the mean HIV/AIDS prevalence rate was 13% for the nine provinces. In other words, it was estimated by way of antenatal testing that on average 13% of a province's total population was infected with HIV. The statistics summarised above are sourced or derived from table 5.1.

Education, income, need and the demand for government services might be contributing factors to the HIV prevalence rate. In 2001 the mean number of individuals with less than a standard 10 education was nearly 872 thousand. The minimum number of individuals with less than a standard 10 education in any one province was a little more than 143 thousand while the maximum number was more than 2 million. On average, close to 19% of a province's population (age 16-65) had income ranging between R400 and R800 rands per month in 2001. Minimum and maximum levels were 6% and 30% respectively. Indeed, in one province 30% of the population had income between R400 and 800 rands per month. In terms of 2005 (USD) dollars, 30% of a provinces population had income of approximately $45 dollars per month – a little more than a dollar a day, or about 7 rands a day. The rationale for first using the dollar (USD) as an indicator of subsistence is because it is a benchmark currency for foreign exchange. Still, for greater clarity, the daily rate of subsistence is reflected in South African rands (about R7/day) to put it into a South African context. For that matter and with reference to need, by 2004 the mean number of individuals not economically active was nearly 1.5 million. These statistics are fairly in line with an unemployment rate that is thought to be around 40%. Since the ANC's assumption of power, the reported jobless rate has ranged from 30% to 35%. These factors, education, income, need (social-welfare factors) and the HIV/AIDS

prevalence rate is presumed to have had some strain [effect] on the demand for government services, as by 2003 the mean change in the demand for provincial government services 5%. Highly skewed, in one province (the minimum) demand for government services contracted by a negative 1%. In yet another province (the maximum) demand for government services increased by 30%. Notably, the provinces' ability to meet the increased demand for services should be reflected in its productivity – i.e., provincial GDP. In 2003, the mean change in GDP was less than 1%. The maximum change in the rate of growth of any one province was a little more than 1%. One province had a change in the rate of growth of a negative five-tenths of one percent – i.e., negative growth. This might perhaps be a feature of the unitary state that, in the case of South Africa, promotes the assignment and transfer of revenue from the central government to sub-national governments [provinces]. Indeed, a significant portion of provincial governments' revenue is derived from national government (Levy & Tapscott, 2001:131). Consequently, the need for provinces to be productive may not have been a necessity, with expectations of revenue deriving from the national sphere of government.

The effects of race are considered firstly by considering the percent of province's population that are white and secondly by the percent of provincial votes received by the New National Party (NNP). In 2001, on average 11% of a province's population was white. One province had a mere 2% of its population being comprised of whites, while another province had 40% of its province being comprised of whites (Statistics South Africa, 2003). In the 2004 elections, the NNP (the remnant of the apartheid National Party) won on average 2% of the popular vote. In one province the NNP was able to win 10% of the popular vote (Independent Electoral Commission, 2004). Surprisingly, by

2006 the NNP was all but non-existent, with its party members being absorbed into the ANC, the Democratic Alliance (DA) and other minor parties.

5.3   BIVARIATE REGRESSION ANALYSIS

There is hesitation to draw final conclusions based on the measures of central tendency presented as descriptives. It, however, would be safe to say that based on the descriptives above provinces are potential breeding grounds for HIV/AIDS. An increase in the prevalence rate is inevitable. In light of an average population growth rate of 6%, provincial spending for HIV/AIDS ranging 11 million to 246 million rands seems hardly enough. Moreover, average conditional grants of R3 million cannot be expected to curtail an HIV/AIDS prevalence rate (on average) of 13%. These statistics were sourced from table 5.1 where descriptive statistics are indicated.

Nevertheless, the descriptives prepares for considering four sets of bivariate relationships. Notably, the relationship between voter turnout and government spending is the primary concern of this dissertation. Can voter's influence government spending for HIV/AIDS? The first set of bivariate models considers the relationship between the dependent variable for provincial spending and the independent variables for voter turnout, the change in population growth and the HIV/AIDS prevalence rate. The set consists of three bivariate models where simple [ordinary] linear regressions (OLS) are run using SPSS. Using the same estimation method, the second set of bivariate models uses the same independent variables but the dependent variable representing national spending for HIV/AIDS replaces provincial spending for HIV/AIDS. In the case of the third set of bivariate models, the dependent variable is again provincial spending for HIV/AIDS but the independent variables of income and need are regressed. Finally, the fourth set of

bivariate models use the same regressed independent variables but the dependent variable in this instance is national spending for HIV/AIDS. Likewise, OLS is the estimation method used for the third and fourth set of bivariate models. Appendices 2, 3, 4, and 5 presents the SPSS output for the various models.

5.3.1 *Provincial Spending and Voter Turnout Models*

The first bivariate model consists of the dependent variable provincial spending for HIV/AIDS being regressed on voter turnout. Notably, the expectation is that there would be a negative relationship between the two variables. In other words, with the hypothesis being that *voters do not have the ability to influence government spending*, it is not expected that an increase in the independent variable would result in an increase in the dependent variable. Beta (the slope of the regression line) then is expected to be negative. The SPSS output summarised in Table 5.2, however, is contrary to expectations.

<u>Table 5.2</u>

<u>**Bivariate Analysis of Provincial Spending for HIV/AIDS**</u>

Parameter Estimates: **First Set**

Dependent Variable = PROV_SPEND_03

Estimation Method: OLS

Independent Variables	Predicted Sign	R	R^2	(000,000) SE	t	sig	Beta
Model 1 - VOTE_TURN_04	Neg	.822	.630	46.1	3.83	.007	.822
Model 2 - POP_GROW_03	Pos	.457	.209	72.0	-1.36	.216	-.457
Model 3 - AIDS_PREV_02	Pos	.482	.233	70.9	1.46	.189	.482

With reference to model 1, the sign for Beta is positive, suggesting a positive relationship between voter turnout and provincial spending. Moreover, the correlation coefficient R, being relatively close to 1, suggests that there is a relatively high correlation (.82) between the two variables. Moreover, there is more than a moderate proportion (.63) of variation explained by the regression model. That is 63% of the variability in government spending is explained by voter turnout. The *t-statistic* indicates that the sample slope is 3.83 (about 4) standard error units below the hypothesized value of zero (0). However, note that the observed significance level of .007 indicates a low level of significance substantiating the rejection of the null hypothesis that there is no linear relationship between the variables. Although not perfectly 1, but nevertheless close to 1 (.82), there is a positive linear relationship between voter turnout and government spending for HIV/AIDS. Finally, the model appears to be a fairly good model for explaining and predicting changes in government spending due to voting activity.

Commenting briefly on models 2 and 3, both models appear to be less than optimal as the correlation coefficients (R) are construed to be closer to zero (0), thereby indicating low correlation to the dependent variables. Moreover, the variables' qualities of predictability are questionable, considering the low proportions of variation that are explained by the respective models. Recognising that in the real world that population growth and the HIV/AIDS prevalence rate have not been motivating factors for government to spend on HIV/AIDS may explain why these variables are poor predictors of government spending. With regards to population growth, it was expected that the sign would be positive – that is, that government would want to spend on HIV/AIDS policy, in response to of a growing population. A similar explanation is offered for predicting that the sign would

be positive for the HIV/AIDS prevalence rate. The Beta for population growth, however, turned out to be negative, indicating a negative relationship between population growth and government spending for HIV/AIDS. Indeed, this has been the case in the real world, suggesting that population growth has not been an influencing factor for spending on HIV/AIDS. Finally, the sign for AIDS prevalence was predicted to be positive and indeed it was computed (the Beta) to be positive. This confirms the real world expectation that in recognition of an increasing HIV/AIDS prevalence rate, the variable would have some relationship to government spending for HIV/AIDS. The models (1 and 2) are discounted, however, due to a low correlation coefficient R and R^2 explaining variability.

5.3.2 National Spending and Voter Turnout Models

Not surprising, there is virtually no correlation (.05) between voter turnout and national spending for HIV/AIDS – Table 5-3. Chapter three of this dissertation recalls the history of national policy making on the matter of HIV/AIDS and the discussion therein supports the output in model 1. A negative sign was indeed predicted, implying that an increase in voter turnout would not result in an increase in government spending. A Beta of .05, although positive, is so close to zero, it more or less compliments the "disconnection" between voters and policy makers that was alluded to in chapter three.

Table 5.3

Bivariate Analysis of National Spending for HIV/AIDS

Parameter Estimates: **Second Set**

Dependent Variable = NATL_SPEND_02

Estimation Method: OLS

Independent Variables	Predicted Sign	R	R^2	SE	t	sig	Beta
Model 1 - VOTE_TURN_04	Neg	.050	.002	2.52	.132	.899	.050
Model 2 - POP_GROW_03	Pos	.498	.248	2.19	-1.52	.173	-.498
Model 3 - AIDS_PREV_02	Pos	.554	.306	2.11	1.76	.122	.554

A hypothesis that voters do not have the ability to influence policy making on HIV/AIDS as reflected in national government spending may be true, with regard for national spending. However, from the start the emphasis was on provincial spending and policy making. Additionally, considering the R^2 explaining variability (and the low correlation) coefficient, the model can only be discounted, offering no predictability of national spending for HIV/AIDS based on voter turnout. There is an inclination to disregard models 2 and 3 as well due to marginal (.49 and .55) correlation between the variables. A decision to do so is also based on the low R^2 explaining variability. Conclusively, models 1, 2, and 3 in the second set are not reliable models and therefore no conclusions can be drawn on the matter of national spending for HIV/AIDS. This process of rejecting models is not uncommon – at least for this study. Characteristically, this is an iterative process. In other words, a number of variables, in some combination, or in some set are paired, regressions run, coefficients calculated and if there are low measures of

variability, than those models should be discounted. The process is repeated, as will be seen until the optimal model is found and variables reflecting relatively significant degree of correlation are realised. A conservative approach is taken in rejecting models that offer little or no explanation for causality.

5.3.3 Provincial Spending Models - Income/Need

The variables income and need are considered in order to account for any income effects upon provincial government spending for HIV/AIDS. The variable NEED_04 is associated with need, as it represents the absence of income, whereas INC_01 at the least represents some [although] minimal amount of income earned. From the data below in the third set, income is not significantly correlated (.50) to provincial government spending. It was expected that the sign would be positive, with government spending targeted at low income earners – i.e., AIDS prevention targeted at the most vulnerable. Indeed, the sign indicated with a Beta of -.504 is negative. The negative correlation between income and provincial government spending is illogical. It should be expected that as the number [percentage] of the population [those] in need of income increases, government would respond with, say, social-welfare programmes. Nevertheless, with only 25% of the variation being explained by independent variable income, the model itself is discounted.

Table 5.4

Bivariate Analysis of Provincial Spending for HIV/AIDS – Need/Income

Parameter Estimates: **Third Set**

Dependent Variable = PROV_SPEND_03

Estimation Method: OLS

Independent Variables	Predicted Sign	R	R^2	(000,000) SE	t	sig	Beta
Model 1 - INC_01	Pos	.504	.254	69.9	-1.54	.166	-.504
Model 2 - NEED_04	Pos	.806	.649	47.9	3.59	.009	.806

A case, however, can be made for model 2 where there is a relatively high (.81) correlation between need and government spending. Moreover, the sign was expected to be positive and indeed it is. There is indication that as there are increasing numbers of citizens who are economically inactive, government responds by spending on, in this case, HIV/AIDS prevention. With an R^2 of .65, a fair proportion of the variability in government spending is explained by the independent variable NEED_04. The model is therefore an adequate model for consideration and further discussion.

5.3.4 National Spending Models – Need/Income

The correlation coefficients (R) in the two models below are close to zero and consequently construed to mean that there are minimal linear relationships. Moreover, the low R^2 indicating strength in explaining variability, as in other instances, renders the models to be inadequate in predicting changes in national government spending for HIV/AIDS. An explanation offered for the inadequacies of the models can be that the data, while adequate for analysing provincial government spending, is inadequate for

analysing national spending. Recall, that the data pertaining to government spending reflects conditional grants allocated to the provinces. Another variable or proxy reflecting national government spending might be used in a follow-on study.

Table 5.5

Bivariate Analysis of National Spending for HIV/AIDS – Need/Income

Parameter Estimates: **Fourth Set**

Dependent Variable = NATL_SPEND_02

Estimation Method: OLS

Independent Variables	Predicted Sign	R	R^2	SE	t	sig	Beta
Model 1 - INC_01	Pos	.073	.005	2.52	-.192	.85	-.504
Model 2 - NEED_04	Pos	.099	.010	2.52	.263	.80	.099

As the focus of this dissertation is on provincial government spending, attention is turned away from the fourth set of bivariate models. Nevertheless, two models are identified to be useful towards explaining or offering causality for provincial government spending.

Table 5.6

Bivariate Models - Provincial Spending HIV/AIDS

Dependent Variable = PROV_SPEND_03

Independent Variables	Predicted Sign	R	R^2	(000,000) SE	t	sig	Beta
Model 1 - VOTE_TURN_04	Neg	.822	.630	46.1	3.83	.007	.822
Model 2 - NEED_04	Pos	.806	.649	47.9	3.59	.009	.806

A least squares regression line can be formulated for each model above and subsequently used to predict values indicating prospective provincial spending for HIV/AIDS – of course dependent upon voter turnout and need (Norusis, 1998:19). In the case of model 1, the *y*-intercept and the slope indicated in Appendix 2 (coefficients) provides for establishing the following predictive [model] straight-line equation:

Equation 5.1

Predicted Provincial HIV/AIDS Expenditure = -35,010,911 + (65.68 x Voter Turnout)

Upon inserting an anticipated level of voter turnout in the model above, an amount indicating provincial expenditure can be predicted. Likewise, a least squares line can be formulated for model 2, upon use of the *y*-intercept and slope indicated in Appendix 4:

Equation 5.2

Predicted Provincial HIV/AIDS Expenditure = -30,043,733 + (72.09 x Need)

The implications of these models will be discussed further, when there is confirmation or rejection of the hypothesis – section 5.6. In the interim, there is merit in being able to predict levels of expenditure for efficient policy making and decision making capability. Of course, as indicated by coefficients in Table 5-6, the efficacy of the models has been empirically shown to be quite adequate.

5.4   MULTIVARIATE REGRESSION ANALYSIS

Thus far two bivariate models have been determined to be adequate predictors of government spending, inclusive of the independent variable voter turnout and

surprisingly the variable encompassing need. The variables pertaining to population growth, HIV/AIDS prevalence and income were discounted due to weak correlations to the dependent variable – especially, when the dependent variable was national spending for HIV/AIDS. Notably, the focus of this dissertation is on provincial government spending but considering the (unitary state) relationship between the national sphere and the provincial spheres of government (revenue sharing), national spending for HIV/AIDS was considered. Bivariate analysis for national spending, at least for the moment here in this dissertation, offered inadequate explanations for causality. Henceforth, multivariate analysis will focus on provincial government spending for HIV/AIDS.

The question now is: What other variables might possibly offer some explanation of causality and influence on the dependent variable PROV_SPEND_03? Can a multivariate model be formulated to reflect the multiplicity of effects, factors, etc. that in reality impact government spending (and so public policy decisions) for HIV/AIDS treatment and prevention? Heretofore, a multivariate model will be tried and tested. A test for multicollinearity will be conducted to eliminate less than optimal variables. Once an optimal model has been found, select variables will be eliminated in order to identify an optimal model to predict or explain provincial government spending for HIV/AIDS.

5.4.1 Multivariate Regression Analysis

Upon reviewing Table 3.2, eight (8) independent variables were identified to be included in the initial multivariate equation. Table 5.7 highlights those variables.

Table 5.7

Variables Included In Multiple Multivariate Equations

	Name	Description
1.	WH_RACE_ 01	Percent of Provincial Population That Are White
2.	EDUCA_01	No. of Individuals w/less than Std.10 Education
3.	LAT_GROUP_04	Latent Group Influence - % Change in Voter Reg. '99 – '04
4.	SPEC_INT_01	Special Interest Group TAC Influence On Aids Policy
5.	Δ_PROV_GDP_03	% Change in Provincial Economic Productivity 2002-2003
6.	Δ_DEM_GOVTSERV_04	Change in Demand for Government Services 2002-2003
7.	NNP_RACE_04	% of Votes Received by New National Party – 2004
8.	NATL_SPEND_02	Conditional Grants To Provinces for HIV/AIDS Spending

Appendix 6 presents the SPSS output to commence analysis to determine an optimal multivariate model. Firstly, backward elimination (Norusis, 1998:470) was used where initially all independent variables were part of the linear regression model and after several steps [recalculations] the variable having the least effect on the model's coefficient of determination (R^2) was subsequently eliminated. Recall that section 3.4.3 discusses multicollinearity and backward elimination. The process of backward elimination is reflected in the following table showing three prospective multivariate models and associated correlation coefficients.

Table 5.8

Summary of Models Resulting From Backward Elimination

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.992[a]	.984	.871	27184077.0
2	.991[b]	.982	.929	20199197.8
3	.991[c]	.982	.952	16550921.1

a. Predictors: (Constant), NNP_RACE_04, WH_RACE_01, LAT_GROUP_04, CHANGE_DEM_GOV_SERV_04, SPEC_INT_01, CHANGE_PROV_GDP_03, EDUCA_01

b. Predictors: (Constant), NNP_RACE_04, WH_RACE_01, LAT_GROUP_04, SPEC_INT_01, CHANGE_PROV_ GDP_03, EDUCA_01

c. Predictors: (Constant), NNP_RACE_04, WH_RACE_01, LAT_GROUP_04, SPEC_INT_01, EDUCA_01

When all seven independent variables (except the variable national spending for HIV/AIDS) are included in the model 1 multivariate model, the correlation coefficient (R) reflecting the overall linear relationship between the independent variables and the dependent variable (provincial government spending for HIV/AIDS) is remarkably high (.99). Through backward elimination certain variables fall by the wayside, due to their minimal affect on R^2. Of course, R^2 indicates the degree of predictability by the independent variables. Model 3 is the resulting optimal model, with the predictors CHANGE_DEM_GOV_SERV_04 and CHANGE_PROV_GDP_03 being eliminated. The linear equation [model] for prediction where the *y*-intercept and betas are shown will be presented in section 5.4.3, Optimising the Multivariate Model.

At this point, a number of comments are warranted. Firstly, the variable NATL_SPEND_02 was immediately and automatically removed – most likely because the observed significance level for the variables coefficient was greater than .10. The process of backward elimination will (to begin) retain those variables where the observed significance level is less than .10. Secondly, it is purely coincidental that (the researcher) was able to identify variables for the multivariate model that contribute near perfectly (.99) to an outstanding linear relationship. The resulting high coefficient of determination is coincidental as well. No explanations other than luck and intuition are offered. Thirdly, note that model 3 and for that matter this initial multivariate analysis does not include the two bivariate independent variables of voter turnout and need. These two variables will be considered shortly hereafter. Finally, partial correlation coefficients for the variables retained in model 3 are as follows:

Table 5.9

Partial Correlation Coefficients for Multivariate Variables

Variable	Partial Coefficient	Beta Sign
NNP_RACE_04	.910	Positive
WH_RACE_01	-.971	Negative
LAT_GROUP_04	-.937	Negative
SPEC_INT_01	-.867	Negative
EDUCA_01	.986	Positive

Whether negative or positive, all the variables show a high correlation to the dependent variable PROV_SPEND_03. The relationships, however, are subject to interpretation. In the case of NNP_RACE_04, the increase in provincial spending relative to an increase in NNP voter activity may be the result of political party activity to influence provincial spending by way of the "white vote." Conversely, an increase in spending resulting from a decrease in the white population may represent [overall] provinces' positive response to a dwindling white voter constituency. Most interesting is the negative signs for latent and special interests groups. As these groups' activities decrease, provincial spending increases. This is an anomaly. It might be necessary to exclude the latent group variable due to multicollinearity (to be discussed in the following section). The negative sign for the special interest group variable may reflect government's unresponsiveness and resistance to the activism of the TAC. Finally, it is somewhat plausible that the variable for education would have a positive sign. That is, as the number of individuals with less than a standard 10 education increases, government responds [theoretically] by spending more, say, on education – e.g., HIV/AIDS prevention education. Again, these explanations are based on interpretation and speculation.

5.4.2   *Multicollinearity*

Section 3.4.3 discussed the possibility of bias in regression coefficients. With regard for multicollinearity, when there is a relationship (linear for example) between predictors or two independent variables, a specification error could occur. Thus, when the tolerance of the variable is close to 1, there is some indication of a linear relationship between two independent variables.   It then becomes necessary to discount a variable from the multivariate equation. Appendix 6 includes a table of coefficients where the tolerances

for model 3 are shown. Notably and as mentioned above, the variable LAT_GROUP_04 has a tolerance of .709. Having the highest tolerance coefficient, a decision is made to exclude the variable from the multivariate [regression] model. There is good cause to exclude the variable from the model, due to the negative sign of the partial correlation coefficient. The expectation is that a latent group's activities would increase government spending. That is, an increase in LAT_GROUP_04 would result in an increase in PROV-SPEND_03. With the latent group variable sign being negative, that expectation has not been met. The high tolerance of .71 further justifies excluding the variable form model 3. Consequently, 4 variables (WH_RACE_01, EDUCA_01, SPEC_INT_01 and NNP_RACE_04) are optimal variables for a predictive model for provincial government spending for HIV/AIDS.

5.4.3 *Optimising the Regression Model*

Bivariate regression analysis was conducted in section 5.3, with the outcome being the identification of two bivariate equations offering cause and effect, and prediction for provincial government spending for HIV/AIDS policy making. The variables for voter turnout and need, on their own, were found to be significant predictors of government spending. Notably, those variables were not considered in the process of multivariate regression analysis. In that instance, select variables were considered and several variables were subsequently excluded by way of backward elimination and scrutiny for multicollinearity [tolerances]. The desire now is to further optimise the model by combining the bivariate variables and the multivariate variables. Appendix 7 shows the regression in full numeric notation for the following six variables:

Table 5.10

Significance of Variables for Optimal Regression

Variable	t	sig
NNP_RACE_04	-.359	.754
WH_RACE_01	-.896	.465
NEED_04	.058	.959
SPEC_INT_01	-1.25	.339
EDUCA_01	3.19	.086
VOTE_TURN_04	-.872	.475

In this instance and when the regression was run, all the variables above entered the regression simultaneously and in one single step. Previously, backward selection was used to eliminate any variable having little or no effect on the models R^2. Again, in this immediate instance all variables enter the regression; arbitrarily, a significance of .50 is used as a cut off to eliminate variables from the regression model. Doing, so, NNP_RACE_04 has a significance of .754 thus preventing rejecting the null hypothesis. In other words, the null hypothesis cannot be rejected, consequently indicating that there is no relationship between the dependent variable and the independent variable. The same can be said of the variable NEED_04 with an observed significance of .95 (Norusis, 1998:410-411). A decision is then taken to build an optimal regression model where the observed level of significance (sig.) is below .50. In each of those cases (WH_RACE_01, SPEC_INT_01, EDUCA_01, and VOTE_TURN_04), the null

hypothesis is rejected. Appendix 8 shows the regression run for those four variables. Note that the observed level of significance (sig.) is below or relatively close to .05 – that level of significance being the usual frame of reference at which a decision is made to reject or accept the null hypothesis. In reference to these variables, a decision is taken to reject the null hypothesis. Conclusively, there is a linear relationship between those variables and the dependent variable PROV_SPEND_03 - be it positive or negative as indicated by the sign of beta. Using the *y*-intercept and betas in appendix 8, the following linear equation is offered as the optimal multivariate regression:

Equation 5.3

Predicted Provincial HIV/AIDS Expenditure =

10,554,928 – (1,131,063,052 x WH_RACE_01) – (141,812,028 x SPEC_INT_01) + (639 x EDUCA_01) – (205 x VOTE_TURN_04)

Although statistically the equation above is considered to be optimal, three of the variables (having negative beta coefficients) will no doubt cause a predicted provincial HIV/AIDS expenditure to be negative. From the model above, it will be concluded that a decrease in the white population, a decrease in special interest activity and, in this case, a decrease in voter turnout results in an increase in government spending. In other words these variables have a negative linear relationship with government spending. A decrease in these variables results in an increase in government spending. With regard for voter turnout, there was a positive linear relationship between the voter turnout variable and the variable for provincial government spending. There is an inclination then, with policy formulation in mind, to run a regression that indeed retains the education variable EDUCA_01 and VOTE_ TURN_04. That regression (Appendix 9) reveals that the

140

variable EDUCA_01 has an observed significance (sig.) of .86 and consequently the null hypothesis cannot be rejected. In other words, accepting the null hypothesis leads to concluding that there is no relationship between EDUCA_01 and PROV_SPEND_03. Thus, as in the initial bivariate regression model, VOTE_TURN_04 is an optimal variable (even here with an observed significance of .47.) that may be used to explain provincial government spending for HIV/AIDS.

It is important to note the effects variables have on R and sig. when running regressions. The correlation coefficient R and the observed significance (sig.) appear to increase or decrease due to the inclusion or exclusion of variables. One explanation offered is the multicollinearity that is detected and subsequently reflected in the tolerances. Having noted that, the regressions that were run yielded three optimal models – two bivariate and the other multivariate. Those models are:

Equation(s) 5.4

Predicted Provincial HIV/AIDS Spending = -35,010,911 + (65.68 **x** Voter Turnout)

Predicted Provincial HIV/AIDS Spending = -30,043,733 + (72.09 **x** Need)

Pred. Prov. HIV/AIDS Spend = -32,159,999 + (20.01 **x** Educa.) + (53.74 **x** Voter Turnout)

An interim conclusion is made that select variables will offer significant explanation for provincial government spending for HIV/AIDS. The variables for voter turnout and need have a bivariate linear relationship with the variable provincial spending. However, in a

multivariate model where voter turnout continues to contribute to predictability, the variable for education also has a linear relationship with the variable for provincial spending. The consequences for policy and inferences will be discussed in section 5.6 where the dissertation hypothesis is either confirmed or refuted.

5.4.4 Test of Hypothesis

Leading to a discussion where the hypothesis is either rejected or confirmed, the discussion now focuses on the bivariate model (equation 5.1) where:

Predicted Provincial HIV/AIDS Expenditure = -35,010,911 + (65.68 x Voter Turnout)

To test the null hypothesis that there is no linear relationship between PROV_SPEND_03 and VOTE_TURN_04 the following table of coefficients is produced:

Table 5.11

SPSS Coefficients for Test of Hypothesis of Slope

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-4E+007	3E+007		-1.061	.324	-113025503	43003680.99
	VOTE_TURN_04	65.679	17.173	.822	3.825	.007	25.072	106.286

a. Dependent Variable: PROV_SPEND_03

The primary concern of this dissertation is for the effect of the voting franchise, as reflected by voter turnout, on provincial government spending. With the correlation coefficient (*R*) being .82, there is good indication of a linear relationship between the dependent variable and independent variable. The relationship is examined by testing the values for the slope (65.679) and the *y*-intercept (-4E+007). Indeed, there is a test of the null hypothesis that the slop is 0. With a small-observed significance (sig.) of .007, the null hypothesis is rejected. Thus, the slope is not zero and there is a liner relationship between the two variables. Furthermore, upper and lower limits for all possible values (95% confidence interval) are indicated in the range of 25.07 and 106.29. Essentially, there are a number of possible values for predicting government spending. It can be said, however, at a confidence level of 95%, certain values will fall between the upper and lower ranges indicated above.

5.5   TEST OF HYPOTHESIS - CORRELATION COEFFICIENT: FISHER'S *Z*

As stated above, a test of hypothesis confirms that the slope (**b**) is not equal to zero – i.e., that there is no linear relationship. With the null hypothesis being rejected, it is accepted that there is a linear relationship. Still, what certainty is there that the correlation coefficient *R* (.82) is truly as calculated the statistic *.82*? *Fisher's z* provides for testing the hypothesis that *R* = *.82*. The test of the [null] hypothesis is written as H_0: **r** = **b**, where in this instance the slope is equivalent to the correlation coefficient *R* because there is indication of linearity and the slope not being zero. The null hypothesis for testing the correlation coefficient can then be written as H_0: **r** = .82 (Kleinbaum & Kupper (1978:78-80), as discussed in section 3.4.6.

To test the null hypothesis, the t-statistic is transformed to a z-statistic through the following formula:

Equation 5.4

$$z = \tfrac{1}{2}\log_e(1 + r/1 - r) - \tfrac{1}{2}\log_e(1 + \boldsymbol{r}/1 - \boldsymbol{r}) \text{ divided by } 1 / \sqrt{n}$$

where:

r = An alternative correlation coefficient in the event that the calculated correlation coefficient is not *.82* – for example, an anticipated lower than expected coefficient of *.68*

$\boldsymbol{r}$ = The calculated correlation coefficient *.82*

The equation is then solved in the following manner - The calculated figures of 0.8291 and 1.1568 are determined from ($\tfrac{1}{2} \ln 1 + r/1 - r$) tables in (Kleinbaum & Kupper (1978:656-657):

$$\tfrac{1}{2}\log_e(1 + r/1 - r) = \tfrac{1}{2}\log_e(1 + .68/1 - .68) = 0.8291$$

$$\tfrac{1}{2}\log_e(1 + \boldsymbol{r}/1 - \boldsymbol{r}) = \tfrac{1}{2}\log_e(1 + .82/1 - .82) = 1.1568$$

$$z = 0.8291 - 1.1568 \text{ divided by } 1 / \sqrt{9}$$

$$z = -1$$

For an alpha (*a*) of .05 where the critical region z is greater than or equal to 1.96, the decision criterion is that of a 95% confidence interval and a one-tail test. Consequently,

with the calculated z being less than 1.96, the null hypothesis cannot be rejected. In other words, the null hypothesis is accepted that indeed the correlation coefficient is .82 and not, say, a different of even a lower coefficient such as .68. Thus, there is additional statistical and empirical evidence of the linear relationship between voter turnout and the ability of the electorate to influence public policy and provincial government spending for HIV/AIDS.

5.6   HYPOTHESIS: CONFIRM v. REFUTE - IMPLICATIONS FOR POLICY

In this dissertation (section 3.2), the hypothesis that voters "do not" have the potential to influence was stated. Chapter 4 could be viewed as having painted a rather dismal state of affairs, as it relates to HIV/AIDS in South Africa – i.e., the inconsistency in policy formulation and implementation by the Mbeki administration.   Consequently, the hypothesis was put forth that the electorate (voter turnout) was virtually powerless, through the ballot box, to impact or influence pubic policy on HIV/AIDS as reflected in provincial government spending for HIV/AIDS. The analysis of the data and bivariate regression where provincial government spending was regressed on voter turnout leads towards concluding that indeed the electorate "can" influence government spending. The stated hypothesis is refuted and rejected. The bivariate regression that was run resulted in a positive linear relationship between the two variables representing voter turnout and provincial government spending for HIV/AIDS. As voter turnout increases, provincial government spending for HIV/AIDS increases. Moreover, there is a relatively high correlation between the two variables, with a fair degree of predictability of the dependent variable by the independent variable. The observed significance level was such that the null hypothesis that the slope of the straight line was zero was rejected.

With regard for policy, there should not be a jump to conclude that government is doing all that it can, in response to the AIDS epidemic. Nor should it be concluded that government is totally responsive to the electorate. It, however, is empirically supported that [provincial] government, to some degree, has been responsive to the electorate. When, as discussed in chapter 4, there is recognition of the revolt against national HIV/AIDS policy by provincial Premiers, it is understandable that at the provincial sphere of government there is empirical evidence of responsiveness. This, unfortunately, is not the case at the national sphere and where national government spending was regressed on voter turnout substantiates national governments unresponsiveness (see section 5.3.2). Conclusively, the hypothesis the voters do not have the ability to influence provincial government spending for HIV/IDS is rejected. The electorate can influence provincial public policy.