CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

CAUSES AND EFFECTS ON PROVINCIAL SPENDING FOR HIV/AIDS

3.1 INTRODUCTION

Before summarising the research methodology, it is important to note that although causal explanations are supported by statistical output, the output does not necessarily prove causation. Nevertheless, the statistical approach and techniques employed do serve to model the world perceived by the researcher and thus enable the researcher to offer a response to the research question based on empirical evidence. Having said that, the research design for this dissertation is evaluative and generally measures a programme. Indeed, that programme is one of HIV/AIDS treatment in South Africa. The design is further qualified by being empirical and quantitative. Furthermore, the research methodology encompasses the use of secondary data, organising and analysing that data through the use of select statistical techniques. Upon identifying numerous variables that might offer some explanation or causal effect on provincial government spending for HIV/AIDS, descriptive statistics will serve to summarise and organise the secondary data, in order that the data can be effectively managed. Thereafter, bivariate and multivariate relationships are established to move towards answering the research question. Indeed, that question is whether the electorate, as reflected by voter turnout, can influence provincial spending on HIV/AIDS treatment. Finally, the hypothesis is that, indeed, the electorate does not have the potential to influence provincial spending and, at then end of the day, such spending reflects a public policy decision.

Thus far this dissertation has entailed, firstly, an introduction where public policy decision-making was presented as the background for this study of the potential for voters to affect or influence provincial government spending for HIV/AIDS. There was motivation in the introduction to focus on the expenditure side of the budget and to focus on that particular line item that represents one of the most pressing policy issues of the day – spending on HIV/AIDS treatment. Secondly, there was a literature review from which select theoretical approaches and concepts will be drawn from to design a statistical model to examine provincial government spending on HIV/AIDS. For example, the notion of a latent group assisting voters to act as a collective (the variable) will be included in the statistical model to be presented in section 3.4 of this chapter.

The emphasis in the model to be presented, however, will be on voter turnout, as suggested in the literature review. However, a number of control variables will be included to control for additional effects on government spending. The task henceforth, after stating the problem and the hypothesis, is to present the statistical approach to be used to examine causes and effects on provincial government spending for HIV/AIDS treatment. Moreover, those variables exhibiting potential causes and effects on government spending will be presented for inclusion in the multivariate equation. This will be followed by a thorough discussion of the statistical techniques to be employed.

3.2 RESEARCH DESIGN

Babbie, Mouton, Vorster and Prozesky (2002:78) offer a schema for determining a research design. Having referred to that work, the research design for this dissertation may be described as empirical, with analysis of existing data. The existing data are numeric, and secondary data that will be used for statistical modelling. Moreover, a quantitative paradigm is applied where there is analysis of select independent variables to determine whether they offer some explanation of causal effects on provincial government spending for HIV/AIDS. A quantitative paradigm, for example, entails assigning numbers as qualitative indicators of the electorate's ability (in the case of this research) to influence a public policy decision by government to budget for and spend on HIV/AIDS treatment. Naturally, an indicator would be voter turnout but a proxy was designed by calculating the change in voter registration (the independent variable LAT_GROUP_04), with the dependent variable for government spending on HIV/AIDS being regressed on the independent variable. Recognising the multi-variability of the construct, the quantitative paradigm also encompasses quantifying the role of a multitude of variables that may describe causes and effects. Finally, the quantitative paradigm requires the researcher to identify sources of error in the research process. This will be evident and discussed shortly hereafter, as there is recognition of inefficiency and biasness resulting from sample size and the sampling approach.

In reference to control variables, where do they [control] variables come from and why are they necessary? Notably, the variables (to be presented and discussed in the section addressing operationalisation) arise from theory (Healy, 1999:431). As it relates to this dissertation, theory is drawn from the literature and is summarized as:

> Control variables that may explain causality and effects on provincial government spending for HIV/AIDS arise from the potential for income effects; the variable income is therefore offered as a possible control variable. Moreover, distributive effects are tested for through the inclusion of a variable reflecting voter turnout, used to test the hypothesis. Collective theory gives cause to consider the potential for latent groups to facilitate voter efficacy - can voters as a collective influence policy decisions? The variable latent group is therefore included in the multiple regression equation.

Indeed, the chapter two literature review offers theories and the basis for considering variables to be included in the multiple regression equation. Those theories are then operationalised as variables, relative to government spending on provincial HIV/AIDS spending. Ultimately, there is a desire to examine the relationship between two or more variables. It, however, becomes necessary to identify what variables should be included, as will be seen and discussed in the section relating to statistical techniques – i.e., test for multicollinearity and test of significance. Notably, Healy (1999:431) emphasised that the world is complex and thus multivariate, even when the discussion is of bivariate relationships. [*sic*]. Thus the statistical approach is to consider many (23) variables and then proceed to identify and use those variables that are most significant.

As there was discussion of a quantitative paradigm being one aspect of the research design, the specific application of statistics is natural in that over time the link between the two (a quantitative paradigm and statistics) has been evident – the link forged and cemented through the work of Galton (1889) in his introduction and subsequent use of the coefficient of regression. Indeed, in keeping with the earliest practitioners of applied statistics through to Stouffer (1950), the subsequent use of statistical analysis is implied in a quantitative paradigm. Notably, Stouffer studied The American Soldier in World

War II Surveys of Men Regardless of Race July-November 1945 (University of Connecticut: Online). Stoufer's work, along with the work of Lerner and Lasswell epitomizes the Policy Science Movement (Cloete & Wissink, 2000:58). Thus, this dissertation remains true to those preceding methodologies and endeavours to apply non-probability sampling as a definitive statistical technique. Essentially, non-probability sampling is non-random sampling. In random sampling, every unit in a population is identified and has an equal chance of being in a [smaller] representative sample reflecting the [larger] population. Such a sampling technique is probability based in that every unit in the population has some chance of being included in a sample. For this dissertation and research, non-probability sampling has been chosen because the sample size is small and not necessarily representative of a population. Plainly, the sample consists merely of nine provinces and does not represent a larger population of provinces. Indeed, numerous provinces or states are not a characteristic of the mid-level subnational sphere of South African Government - unlike the United States where that are 50 states [provinces]. In that instance a representative and random sample could be generated from the population.

At best, a non-probability approach offers convenience – convenience sampling being non-probability sampling. Indeed, convenience sampling chooses, say, sample participants based on the relative ease of access. Characteristically, the sample is self-selected. The convenience, in the case of this research and dissertation is that of not having to draw a sample from a population. This is in contrast, for example, to drawing a sample from the 284 municipalities in South Africa. Random sampling would be applied in this instance and probability sampling used to examine municipal spending on

HIV/AIDS treatment programmes - government spending. This, however, is not the focus of this dissertation but this research can serve as a foundation for examining municipal spending for HIV/AIDS. Notably, convenience sampling does not produce a fair representation of a population. But this is not a problem for this research, as the aim is not to draw back to a population of provinces, as none exits. There are merely a small number of cases (9) that are readily available for use. Thus it becomes necessary to draw attention to *caveats* associated with small a sample size and non-probability sampling.

Because the research associated with this dissertation focuses on the nine provinces in South Africa, the sample size cannot help but to be small. The study might be redirected to examine government spending on HIV/AIDS at the local government or municipal sphere of government but an assumption is made that most municipalities (with exception to the largest metropolitan cities) are not engaged in spending for HIV/AIDS. A budget review of those 284 municipalities would be in order but time and resources do not allow for such a review. Indeed, this could be the basis for a follow-on study. Having said that and recognising the small size of the provincial sample, undoubtedly questions must be raised as to the efficiency of any estimator (Healy, 1999:154-156). A small sample (N_1) will yield a standard deviation (s) that will be higher relative to, for example, a larger (N_2) sample. Indeed, the efficiency of the estimator is relative to the size of the sample. The larger the sample, the more reliable the estimator – e.g., the standard deviation (s). Thus there will be a requirement to critique the efficiency of any estimator generated for the [provincial] sample. Secondly and with regard to non-probability sampling, a question is raised as to the representative-ness of the sample and the ability to make inferences about the greater population. Since there is no population (only nine provinces in total), this caveat is extended to whether any conclusion or inferences can be made and applied to explain causes for and effects on government spending for any one province, or the total sum of all the provinces. In short, accuracy of statistical significance will be questionable. Finally, ideally an estimator that is unbiased is most desirable, thus giving greater creditability to any sample statistic – when the statistic is unbiased, there can be certainty that the estimator is representative of the population, or in the case of this research, the total sum of the provinces. It can then be expected that any estimation of the parameter mu (m) would therefore be accurate. Due to the small sample size and non-random sampling, statistical outputs must be scrutinised for efficiency and biasness.

In defence of a small sample size and non-random sampling, using small samples has been used to study processes that are common to groups – say, a group of people. Moreover, where non-random sampling has been used, the absence of random selection may be offset by the accuracy of the basic [input] data. Truly, in this dissertation secondary data are being used and (as noted earlier) cause has been given to question the efficacy of data collecting organisations (Statistics South Africa). An assumption, however, is made that for the most part secondary data used here has depth, relevancy and are relatively accurate. Certainly, the ability of government and quasi-government organisations in the collection of [census] data will be more reliable than any one individual's efforts. Nevertheless, caution is called for while relying on a small sample and using non-random sampling. Employing such caution, Chow (2000) showed and proved that convenience sampling need not detract from generality in findings. Research designs and methodologies that (probability sampling and the like) yield ambiguous results should be subject to scrutiny. [*sic*] Moreover, small samples have been used in

Vallecillos and Moreno (2002) - albeit a relatively small sample of 49 students in an effort to study learning of statistical inference. Convenience sampling has been used in Dunn and Horgas (2004). Non-probability sampling features prominently in Mccammon (1994) and Scott (1974). Indeed, representivity can be achieved in the use of small samples and non-random sampling and the results have the potential to be unambiguous. Notably, those works cited have been consulted to guide in determining the appropriate statistical technique to use in order to minimise bias, inefficiency, and ambiguity in data output and analysis.

Conclusively, the research design is summarised as being characteristically evaluation based in that there is measurement of a programme (some might say lack of a programme) for HIV/AIDS treatment (Babbie *et al.*, 2002:355). Being evaluation based, an empirical-quantitative paradigm is applied. Thus a quantitative paradigm leads to applying definitive statistical techniques - i.e., to be discussed hereafter. The sample size is small and non-probability sampling will be used, thus implying limitations of the statistical output. Finally, the evaluative empirical-quantitative statistical research design is augmented by a limited qualitative research design entailing a case study touching on being longitudinal in that a historical time line (figuratively) is drawn to depict the linear progression (denial to acceptance) associated with the actualisation of treatment for HIV/AIDS in South Africa (Babbie *et al.*, 2002:398).

3.3 OPERATIONALISATION: METHODOLOGY

To begin operationalising the research methodology, there is reference to ProDEC (Babbie *et al.* 2002:72) to reiterate (1) the research problem, (2) the research design (3) the quest for empirical evidence and (4) the need to draw conclusions. Firstly, the research problem is the cautious but slow pace at which the government of South Africa has responded to the AIDS epidemic. Initially, the response had been one of denial. Over time there has been recognition that indeed HIV causes AIDS but government spending on HIV/AIDS treatment programmes has not been optimal. This linear progression of denial to acceptance will be discussed in the following chapter. Notably, HIV/AIDS spending will be the dependent variable and the research question is: can voters influence public policy decision making on the mater of spending on HIV/AIDS treatment programmes? In this instance the independent variable will be voter turnout. Secondly, the research design was discussed in the preceding section but is again stated to be an evaluative, empirically quantitative, and a qualitative longitudinal case study. Thirdly, empirical evidence will answer the research question supporting either a response in the affirmative, a response in the negative or a suggestion of inconclusiveness. The empirical evidence will be in the form of the statistical output resulting from applied statistical techniques. Those techniques will be discussed shortly hereafter. Fourthly, this dissertation and research will conclude (the final chapter) by drawing appropriate conclusions. As such, ProDEC serves as a guide in outlining the problem, the research design, the collection of evidence and arriving at a conclusion.

Understanding the research problem is key to this dissertation. Achieving that understanding is an objective of the case study but for the moment the problem stems from, first, denial and inaction on HIV/AIDS treatment in South Africa, manifested by inadequate government spending for HIV/AIDS treatment. Over time, denial and inaction have given way to a public policy on HIV/AIDS treatment in South but the policy still does not reflect a sense of urgency in response to the epidemic. Can constituents by exercising their franchise to vote influence public policy decisions to be made by government on HIV/AIDS spending? The research hypothesis now stated is:

Voters do not have the potential to influence public policy decisions by exercising their franchise to vote.

The specific public policy decision is provincial governments' deciding to spend on HIV/AIDS treatment; such spending (amounts expended) is a reflection of a public policy decision. Notably, the unit of analysis entails a social intervention – i.e., spending, implementing an HIV/AIDS treatment programme, or the policy to combat the negative effects of HIV/AIDS. Thus the unit of analysis (a social intervention) is considered to be a *world one object*, with the characteristic of being a real life endeavour lending itself to empirical research (Babbie *et al.* 2002:84-85). This unit of analysis is essentially an action or decision structured to achieve definite goals and objectives. Whether those goals and objectives have been achieved remains to be confirmed or refuted in the chapter involving the analysis of data output and subsequent drawing of conclusions.

The first steps towards operationalising the research were taken by using a non-random sample and by conforming to using a small sample size. That sample (the cases) consists of the following nine provinces:

Table 3.1

Individual Cases (Provinces) Comprising the Sample
1. Western Cape
2. Free State
3. Gauteng
4. Northern Cape
5. KwaZulu Natal
6. North West
7. Eastern Cape
8. Limpopo
9. Mpumalanga

As a start, the research entails examining causes and effects on government spending for HIV/AIDS by each of the provinces indicated above. The primary causal effect (bivariate relationship) is between provincial government spending on HIV/AIDS and provincial voting. Babbie *et al.* (2002:81) noted that in order to show a causal relationship exist between two variables, there is a requirement that the cause precedes the effect in time. Therefore, certain variables should characteristically predate, say, government spending on HIV/AIDS. To that end, the following are independent variables to be included in the multivariate equation, with those variables suffixed by two numbers indicating the year (time element) leading up to the effect on provincial government spending.

Table 3.2

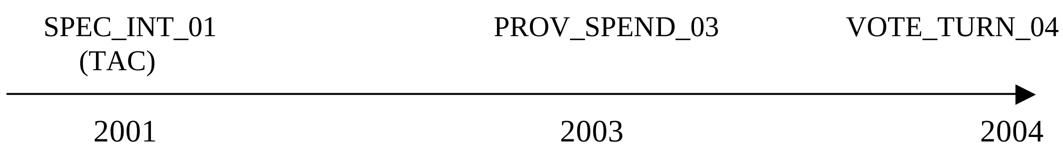
Variables Included In Multiple Multivariate Equations

Name	Description
1. VOTE_TURN_04	Provincial Voter Turnout – 2004 Provincial Elections
2. PARTY_EFF_04	Party Effects/Dichotomous Var. (1) ANC Prov. Legislature
3. WH_RACE_ 01	Percent of Provincial Population That Are White
4. EDUCA_01	No. of Individuals w/less than Std.10 Education
5. INC_01	% of Prov. Pop. (Age 15-65) w/income R400-800/month
6. AIDS_PREV_02	Prov. HIV/AIDS Prevalence Rate at July 2002
7. NEED_04	Those Not Economically Active at March 2004
8. LAT_GROUP_04	Latent Group Influence - % Change in Voter Reg. '99 –'04
9. SPEC_INT_01	Special Interest Group TAC Influence On Aids Policy
10. Δ_PROV_GDP_03	% Change in Prov. Economic Productivity 2002-2003
11. PROV_SPEND_03	Provincial HIV/AIDS Expenditure for 2003
12. REG_VOTERS_99	Registered Voters for 1999 Elections
13. REG_VOTERS_04	Registered Voters for 2004 Elections
14. NATL_SPEND_02	Conditional Grants To Provinces for HIV/AIDS Spending
15. TOTAL_POP_01	Total Provincial Population 2001 Age 15 - 65
16. TOTAL_POP_03	Total Provincial Population 2003 Age 15 - 65
17. POP_GROW_03	Growth Rate for years 2002-2003
18. PROV_GDP_02	Provincial GDP for 2002 – Provincial Economic Growth
19. PROV_GDP_03	Provincial GDP for 2003 – Provincial Economic Growth
20. DEM_GOVTSERV_02	Demand for Government Services - 2002
21. DEM_GOVTSERV_03	Demand for Government Services - 2003
22. Δ_DEM_GOVTSERV_04	Change in Demand for Government Services 2002-2003
23. NNP_RACE_04	% of Votes Received by New National Party – 2004

Again, the longitudinal time factor of each variable is represented by the last two numbers that indicate the year of the inception of the causal effect on provincial government spending on HIV/AIDS treatment. A representative time line is presented as follows:

<u>Figure 3.1</u>

Time Line Illustrating Causal Relationships



The aim above is to show that the cause precedes the effect. Certainly, in the case of activism by the Treatment Action Campaign, the cause-effect relationship would be evident. The notion of cause preceding effect, however, is not absolute. The hypothesis (construed) is that government spending will not increase due to voter turnout but clearly voter turnout above follows government spending. This should not be surprising as, theoretically, the party of the day would be expected to accelerate government spending prior to elections in an effort to acquire votes, fulfil prior campaign promises or to facilitate their remaining in power. For the most part, the 23 variables listed may be placed somewhere along the time line shown above in an examination of a causal relationship with the dependent variable PROV_SPEND_03. Thus there is an attempt to show longitude, in that there is a time factor attached to variables that hypothetically influence, or to be determined, will not influence the dependent variable. Whether this is true one way or another remains to be determined. Conclusions, however, will be subject to the statistical techniques used and subsequent data analysis.

3.4 VARIABLES AND DATA SETS

Tables 3.3 through to 3.7 shows the sample classes, variables and associated data sets. The attached annexure [appendix] provides a reference list for the following data sets.

Table 3.3

Variables and Data Sets

PROVINCE	VOTE_TURN_04	PARTY_EFF_01	WH_RACE_01	EDUCA_01	INC_01
W. Cape	1,566,949	0	.19	1,038,110	.07
Free State	1,011,606	1	.06	482,224	.30
Gauteng	3,408,308	1	.41	2,055,855	.06
Northern Cape	318,702	1	.02	145,344	.25
KwaZulu Natal	2,741,265	0	.11	1,447,674	.15
North West	1,298,563	1	.06	619,263	.17
Eastern Cape	2,231,543	1	.07	963,428	.19
Limpopo	1,614,514	1	.03	653,487	.28
Mpumalanga	1,11,692	1	.05	440,640	.21

Table 3.4

Variables and Data Sets

PROVINCE	AIDS_PREV_02	NEED_04	LAT_GROUP_04	SPEC_INT_01	Δ_PROV_GDP_03
W. Cape	.04	1,029,000	.25	.49	-.19
Free State	.17	811,000	.08	.19	-.54
Gauteng	.16	2,047,000	.13	.06	-.37
Northern Cape	.08	250,000	.18	.00	.21
KwaZulu Natal	.18	2,862,000	.15	.09	-.06
North West	.15	1,167,000	.19	.01	1.35
Eastern Cape	.11	2,331,000	.41	.00	.61
Limpopo	.11	1,918,000	.18	.17	-.35
Mpumalanga	.17	908,000	.14	.00	-.23

Table 3.5

Variables and Data Sets

PROVINCE	PROV_SPEND_03	REG_VOTERS_99	REG_VOTERS_04	NAT_SPEND_02	TOT_POP_01
W. Cape	54,300,000	1,776,021	2,220,283	2.23	4,524,335
Free State	34,800,000	1,225,306	1,321,195	3.48	2,708,775
Gauteng	155,300,000	4,119,164	4,650,594	2.65	8,837,178
Northern Cape	11,300,000	368,205	433,591	.23	822,727
KwaZulu Natal	246,500,000	3,309,162	3,819,864	4.68	9,426,017
North West	42,900,000	1,465,298	1,749,529	8.43	3,669,349
Eastern Cape	70,900,000	2,024,409	2,849,486	1.17	6,436,763
Limpopo	41,700,000	1,858,509	2,187,912	2.94	5,273,642
Mpumalanga	32,300,000	1,266,938	1,442,472	4.20	3,122,990

Table 3.6

Variables and Data Sets

PROVINCE	TOT_POP_03	POP_GROW_03	PROV_GDP_02	PROV_GDP_03	DEM_GOVSERV_02
W. Cape	4,615,965	.02	.0414	.0335	.0178
Free State	2,931,662	.08	.0396	.0184	.0230
Gauteng	9,142,158	.03	.0487	.0307	.0730
Northern Cape	1,011,774	.19	.0146	.0177	.0219
KwaZulu Natal	9,556,833	.01	.0278	.0262	.0026
North West	3,906,592	.06	.0164	.0386	.0001
Eastern Cape	7,244,554	.11	.0166	.0268	.0040
Limpopo	5,535,670	.05	.0420	.0274	-.0094
Mpumalanga	3,160,127	.01	.0275	.0211	.0005

Table 3.7

Variables and Data Sets

PROVINCE	DEM_GOVSERV_03	Δ_DEM_GOVSERV_04	NNP_RACE_04
W. Cape	.0131	-.2640	.1088
Free State	.0780	2.3913	.0082
Gauteng	.0159	-.7822	.0076
Northern Cape	.0007	-.9680	.0752
KwaZulu Natal	.0037	.4231	.0052
North West	.0036	35.0000	.0043
Eastern Cape	.0013	-.6750	.0063
Limpopo	.0062	-1.6596	.0046
Mpumalanga	.0079	14.8000	.0046

Distributive effects are tested for by regressing the dependent variable provincial government spending for HIV/AIDS on the variable representing voter turnout and other independent variables. The variable voter turnout reflects the electorate's participation in the 2004 provincial election; notably, this variable is the primary independent variable. An electorate-voting pattern is reflected in the percentage of votes received by the New National Party that traditionally reflects the white vote and white voting patterns. The potential for race being a causal effect is further tested by representing the percent of a province's population that is white. Not only reflecting a voting pattern but reflecting collective activism as well, the variable latent group represents a facilitating institution's efforts to mobilise voters [constituents] to influence policy decisions. This is reflected in the increase in registered voters for the years 1999-2004. In addition to latent effects, the effect of a special interest group's activism is represented by the number of provincial representatives attending the Treatment Action Campaign's annual conference. As discussed earlier, income effect is tested with the use of the variable income and is extended to include need, as reflected by that variable indicating those not economically active. The variable representing the change in the demand for government services is introduced to test for any effect on government spending for HIV/AIDS due to increase demand for government services. That variable is but one of several variables introduced to explain the variability associated with social problems and issues, as discussed earlier. Other variables include education, the widespread prevalence of HIV/AIDS in a province, population growth, national spending determined by conditional grants from the national sphere of government, political party effects on government spending resulting from control of the provincial government by the ANC and growth of the provincial economy. Again, such variables are introduced to test for additional causes and effects for provincial government spending on HIV/AIDS treatment programmes. Notably, a number of the variables may eventually be eliminated upon testing for multicollinearity. Moreover, upon a test of significance [t-test], some variables may be found to be statistically insignificant. This leads to further discussion of the statistical approach to be used to test the hypothesis that voters do not have the ability to influence public policy on spending on HIV/AIDS treatment programmes.

3.5 STATISTICAL APPROACH AND TECHNIQUES

The statistical approach for this dissertation first concentrates on the bivariate relationship and then look potential multivariate relationship between the dependent variable or provincial government spending for HIV/AIDS and those variables determined to be unbiased and efficient predictors. The statistical techniques that will be used include simple straightforward calculations encompassing descriptive statistics, followed by calculating the *beta* [slope] for simple linear equation. Thereafter, a number of additional independent variables will be considered but several will be eliminated through a test for multicollinearity. Once, the most efficient predictors have been identified, multivariate analysis will be undertaken and a multiple regression [model] will be developed to account for added causes and effect on the dependent variable. During the course of multivariate analysis, a test of significance will further eliminate those independent variables [predictors] that are insignificant and offer little or no explanatory effects on the dependent variable. Finally, a test of hypothesis will be conducted to determine if the beta calculated in the bivariate linear equation is truly representative of the statistical outcome.

3.5.1 Descriptive Statistics

To begin analysing the data sets shown previously, descriptive statistics will be used to summarise and organise the data into forms that will facilitate immediate understanding. A measure of central tendency that will be calculated, for example, includes the arithmetic mean. Measures of variability that will be calculated include the standard deviation and the coefficient of variation. The mean will indicate the average for a particular variable. Once the mean has been found, the standard deviation will be calculated to calculate the distance of the scores (a piece of data) from the [mean] measure of central tendency. In other words, there is a calculation of the dispersion of data around a particular mean. The coefficient of variation will facilitate an analysis of the variability of variables. The coefficient of variation expresses the standard deviation as a percentage of the mean. Finally, a standard score will be calculated to indicate the number of standard deviations a case is above or below the mean. This provides an additional reference point (the first being the mean) that will enable a unit of data to be compared to yet another unit of data

3.5.2 Bivariate Regression & The Coefficient of Determination

Following the elimination of those variables [predictors] that have the highest correlation relative to other predictors, and following the computation of descriptive statistics to organise data, a simple regression model or bivariate regression will be presented to estimate bivariate regression coefficients (Stiefel, 1990:13). This simple regression model goes right to the heart of the hypothesis in that it will examine the relationship between the dependent variable (government spending for HIV/AIDS) and the independent variable (voter turnout). Generally, this is represented by the regression equation $Y_i = b_0 + (b_1 . X_i) + e_i$ where:

Y_i = Provincial Government Spending for HIV/AIDS Treatment

X_i = Provincial Voter Turnout

b_1 = The Slope of the Regression Line and Represents the Change in *Y* divided by the Change in *X*

b_0 = The Intercept or Where the [Regression] Line of Best Fit Cuts Across the Y axis

e_i = An Error Term for Randomness and the Stochastic Relationship of Inefficient Predictors

Basically, the ordinary least squares method is applied and the sum of the squared errors is minimised. Notably, minimising the sum of the squared errors and solving for the associated normal equation is done to obtain estimators for the regression line. The two normal equations referred to are: $\Sigma(Y - b_0 - b_1 . X) = 0$ & $\Sigma(Y - b_0 - b_1 . X)(X) = 0$. The first equation derives the estimator for the Y intercept and the second equation derives the estimator for the slope (Steifel, 1990:22). Thus, the transformed equation for the *Y*-intercept is $b_0 = Y_{[bar]} - b_1 . X_{[bar]}$ and the transformed equation for the slope is $b_1 = (\Sigma xy) / \Sigma x^2$). Consequently, estimates of actual values for independent variables can be obtained and a value for the (*Y*) dependent variable can be calculated. A question arises, however, as to the appropriateness of the regression line. Is it a good predictor of possible outcomes? For that reason, the coefficient of determination will be calculated.

The coefficient of determination is calculated by dividing the explained variation by the total variation. Total variation is the $\Sigma(Y_t - Y_{[bar]})^2$ - i.e., the difference between the actual value of *Y* and the mean of *Y*, squared and totalled for all observations. Explained variation is $\Sigma(Y_{t[hat]} - Y_{[bar]})^2$ - i.e., the difference between the value of each predicted *Y* along the regression line ($Y_{t[hat]}$) and the mean of *Y* , squared and totalled for all observations. Thus the coefficient of determination $R^2 = \Sigma(Y_{t[hat]} - Y_{[bar]})^2 / \Sigma(Y_t - Y_{[bar]})^2$. The calculated coefficient will range between 0 and 1. A high value will value close to 1 will indicate high degree of explained variation. Notably, the coefficient of determination is the equivalent of Pearson's (*r*) coefficient squared and likewise indicates the degree of association between the two variables. Values close to 0 indicate little linear or no linear association while a value close to one indicates a strong linear association between the dependent variable government spending for HIV/AIDS and the independent variable voter turnout (Healy, 1999:394).

3.5.3 Multicollinearity

Freund and Minton (1979:92-93, 112)) alerted the researcher that bias in regression coefficients can result from inadequate specification. Indeed, a *specification error* may occur when using too many variables and some of the variables are truly irrelevant. Thus a measure to be used towards minimising bias and inefficiency is to test for multicollinearity and eliminate those independent variables that are [statistically] shown to be irrelevant. Notably, multicollinearity exists when there is a correlation amongst predictors. Consequently, two steps will be taken to eliminate those variables that are irrelevant.

Firstly, a tolerance test or a test of the linear relationships amongst independent variables will be conducted. Notably, the test for tolerance involves calculating a proportion (Norusis, 1998:467-468). Values calculated, therefore, range from 0 to 1. The closer the variable is to 1, the more certainty there is that the variability in independent variable is explained by another independent variable. Conversely, the closer the variable is to 0, the more certainty there is that the independent variable is closely associated (has a relationship) with some other independent variable. By running the SPSS menu item *Collinearity Diagnostics,* the indicator will be calculated. Multicollinearity indicators for all prospective variables can be ranked, choosing those that are closet to the value 1.

Secondly, a secondary *backward elimination* (Freund & Minton, 1979:22) approach will be used where the t-statistic for each coefficient is determined. Thereafter, the regression coefficient with the minimum absolute t-values will be eliminated. Once a number of statistically insignificant variables are eliminated, the final result will be an optimal model to explain causal effects in provincial government spending for HIV/AIDS.

3.5.4 Multivariate Analysis

Howell (1989:134) stressed the appropriateness in asking: how well some linear combination of two, three, even four predictors (independent variables) influence a dependent variable. Indeed, there is no reason to limit the regression equation to the bivariate form. Table 3.2 offers numerous variables that offer explanation (some more than others) – variables having some perceived causal effect or linear relationship with the dependent variable provincial government spending for HIV/AIDS. The objective of multivariate analysis is to observe the effect of other variables on a bivariate relationship. Notably, the bivariate relationship was indicated on the preceding page. After specifying the bivariate relationship, the objective is to measure effects of other significant variables (Healy, 1999:417). Thus, certain additional variables are fixed i.e., their value no longer free to vary. Importantly, the impact of any bivariate relationship can then be assessed. It, therefore, is natural that multivariate analyses follow bivariate analysis in order to acquire a greater understanding of the relationship between government spending and voter turnout. This, however, will take place upon eliminating, by testing for multicollinearity, those variables that are biased and inefficient.

The following multiple regression formula is offered as a starting point for describing the overall linear relationship between the dependent variable and multiple independent variables found to be the most efficient predictors. That multiple regression equation is (Healy, 1999:448):

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

where:

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

a = The Y-Intercept

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

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

Notably, the coefficients ($b_{1,}$ $b_{2,}$ b_3 . . . $b_{n)}$ indicate partial slopes and represent the amount of change in *Y* for a unit of change in *X*. Importantly, effects of other independent variables in the equation will be taken into consideration. Indeed, the betas are partial coefficients of correlation that represent the effect of the associated independent variable on the dependent provincial variable government spending for HIV/AIDS.

3.5.5 Multiple Correlation and the Coefficient of Multiple Correlation

Once the linear relationship between each independent variable and the dependent variable has been established, the combined effects of all the independent variables will be determined by calculating the coefficient of multiple correlation (R^2). In other words, taking into consideration all of the variables in the multiple regression equation, to what extent [simultaneously] do all variables collectively explain the proportion of variance in the dependent variable? With the correlation coefficient being represented by *R*, as seen above with *Pearson's r*, in the case of multivariate analysis the following formula is offered to calculate to calculate R^2:

$$R^2 = r^2_{y1} + r^2_{y2.1}\,(1 - r^2_{y1})$$

where:

R^2 = The Multiple Correlation Coefficient

r^2_{y1} = The Zero-Order Correlation Between the *Y* and X_1 Variables With the Quantity Squared

$r^2_{y2.1}$ = The Partial Correlation of the *Y* and $X_{2,}$ While Controlling for X_1 With the Quantity Squared

The first term r^2_{y1} is the coefficient of determination for the relationship between the dependent variable and, say, the first independent variable. Indeed, it represents the amount of variation in the dependent variable explained by that particular independent variable. Added to the amount of r^2_{y1} is an amount that represents additional explained variation due to a second dependent variable, as represented by $(1 - r^2_{y1})$. Notably, $r^2_{y2.1}$ provides for controlling for the effects of the first dependent variable. In this instance, the first independent variable is construed to be the primary independent variable provincial government spending for HIV/AIDS; the second independent variable could be any secondary efficient and unbiased predictor that is included in the multiple regression equation indicated above in section 3.4.4. Consequently, the coefficient of determination allows for evaluating the combined explanatory effects of, in this case, two independent variables on the dependent variable and serves to strengthen the information gained through having first examined the [primary] bivariate relationship (Healy, 1999: 417). Importantly, before solving for R^2 it is necessary to calculate r that represents partial correlation of Y and X_2 (the second independent variable), while controlling for $X_{1.}$ The formula for partial correlation where (r_{12}) equals the bivariate correlation between X_1 and X_2 (Healy, 1999:445-449).is:

$$r_{y2.1} = r_{y2} - (r_{y1})(r_{12}) \, / \sqrt{(1 - r^2_{y1})} \; \sqrt{(1 - r^2_{12})}$$

Source Healy, 1999:456

3.5.6 Test of Hypothesis and Confidence Intervals for r

The relationship between the dependent variable and the independent variable can be further tested by a test of the null hypothesis H_0: $r_1 = 0$. In the case of this dissertation, the dependent variable is government spending for HIV/AIDS and the independent variable is voter turnout. After calculating a beta (b_1) through the process of bivariate regression, a subsequent question arises as to whether the calculated beta represents a true correlation between the dependent variable and the independent variable (Kleinbaum & Kupper, 1978:79). When b_1 is calculated, a figure close to 1 indicates a strong linear association between the dependent variable and the independent variable. The hypothesis here is that the electorate (reflected by voter turnout) does not have the ability to influence provincial government spending. It, therefore, is expected that b_1 is close to zero, possibly some negative number. Once b_1 has been determined, the question arises as to: how reliable is that particular coefficient as a predictor? Notably, a test of hypothesis for the efficiency of b_1 is analogous to a test of hypothesis for *r*. In other words, H_0: $r = 0$ is equal to H_0: $b = 0$ (Kleinbaum & Kupper,1978:58-59, 88).

With a test of the null hypothesis for b_1 by way of testing for r_1 , there is a possibility that the distribution will be skewed. Consequently, *Fisher's Z transformation* is used to set confidence limits for testing the hypothesis that H_0: $b_1 = 0$. Fischer's transformation encompasses a log transformation as indicated in the flowing formula:

$$½ \log_e(1 + r / 1 - r) \pm z_{1 - a / 2} / \sqrt{n}$$

In the equation above, $z_1 - \alpha / 2$ provides for a two-tail test to establish lower and upper limits at some particular confidence interval – i.e., 95% or 99% confidence intervals. The log transformation (log_e) provides for the instance that a normal distribution is not evident. Notably, logarithmic transformation tables (Kleinbaum, & Kupper, 1978: 656-657) for ½ log_e ($1 + r / 1 - r$) are used to determine the upper and lower limits used to reject or accept the null hypothesis that beta (b_1) represents a true correlation, or as hoped in the case of this dissertation that beta substantiates the hypothesis that voters can not influence public policy decision with regard to provincial spending for HIV/AIDS. Essentially, the null hypothesis is H_0: $r = b_1$ and that beta is representative of the calculated outcome. Alternatively, the alternative hypothesis would be H_A: r ? b_1, indicating that the calculated beta is not a representative coefficient.

3.6 SUMMARY

Notably, the statistical techniques [calculations] indicated in the preceding sections will be accomplished using SPSS. The formulas, however, provide a foundation for understanding the data output and subsequent analysis. This chapter entailing the research methodology has discussed the research design. Indeed, the design is described as empirical analysis of secondary data pertaining to provincial government spending for HIV/AIDS treatment and voter turnout. A simple bivariate model will determine the relationship where the slope (b) will reveal, say, a linear or non-linear relationship between the independent variable and the dependent variable. Through multivariate analysis, the influence of other variables on government spending will be taken into consideration. Whatever the outcome, caution is stressed. Admittedly, the sample size is small and the statistical approach is non-probabilistic. Notably, convenience sampling is part of the research design. As stipulated early on, causal explanations may be supported by the data output and subsequent analysis (chapter 5) but whether the data is proof of causation is debatable. Before analysing the data it would be appropriate to, by a case study of perspectives, examine policy making for HIV/AIDS treatment. Where possible, rational choice theory (chapter 1) will be used as a framework for discussion.