

A Practical Guide to the Use of Selected Multivariate Statistics

This document is set up to allow the user of multivariate statistics to get assistance in the choice of multivariate technique to use and the state the data must be in to use the desired technique. A series of links enable the user to view areas of interest in relation to multivariate statistical theory, use, and application.

Cette information aidera l'utilisateur des statistiques multivariées à choisir la technique appropriée et à traiter correctement les données pour qu'elles soient dans l'état requis, afin d'utiliser la technique voulue. Divers liens permettent à l'utilisateur d'obtenir de l'information sur la théorie, l'utilisation et l'application des statistiques multivariées.

Les scientifiques ayant participé au projet étant anglophones, Guide pratique d'utilisation de certaines techniques en statistiques multivariées est offert seulement en anglais.

Mike Wulder

Research Scientist
Canadian Forest Service
Pacific Forestry Centre
506 West Burnside Road
Victoria, BC V8Z 1M5
Canada

Main Index:

| | |
|--|----|
| 1. Introduction to this Document | 2 |
| 2. Summary of Multivariate Statistics | 2 |
| 3. Data Screening | 3 |
| 4. Multiple Correlation and Regression | 4 |
| 5. Principal Components Analysis and Factor Analysis | 7 |
| 6. Discriminant Analysis | 12 |
| 7. Cluster Analysis | 15 |
| 8. Spatial Autocorrelation | 17 |

Appendixes:

| | |
|---|----|
| 1. Census data. | 19 |
| 2. Correlation Coefficient and Coefficient of Variation | 39 |
| 3. Homoscedasticity | 40 |
| 4. Linearity | 41 |
| 5. Multivariate General Linear Hypothesis (MGLH) | 42 |
| 6. Missing Data | 43 |
| 7. Multicollinearity and singularity | 43 |
| 8. Normality | 44 |
| 9. Orthogonality | 46 |
| 10. Outliers | 46 |
| 11. Residual Plots | 48 |
| 12. Data Transformations | 50 |
| 13. References | 56 |

1: Introduction to this document

The objective of Graduate Geography 616, Multivariate Statistics, with Dr. Barry Boots, was to familiarize the participant with a range of multivariate statistical procedures and their geographical applications. This was accomplished in two ways, through lectures which provided an introduction to the underlying theories, and through subsequent application of the theory on a typical geographic data set. Stress was placed upon the limitations and validity of individual procedures in sample empirical contexts and the ability to interpret the results of any particular analysis. The multivariate statistical techniques examined were all applications of the multivariate general linear hypothesis. The four principle techniques which were examined, and which will be assessed, are:

1. multiple correlation and regression,
2. principal components and factor analysis,
3. discriminant analysis, and
4. cluster analysis.

To properly apply the general linear model the data input to such analysis must meet specific constraints and criteria, such as the rules of normality, linearity, homoscedasticity, and non-multicollinearity, to accomplish this, a guide to data screening is presented. Spatial autocorrelation also affects many geographic data sets and may need to be addressed.

The goal of this document is to present a summary of the statistical techniques presented and to demonstrate each with examples and illustrations. Consult the references list for depth greater than appropriate from a document of this scope. The data set which is used for this study is an excerpt of Canada Census data for 1991 (see appendix 1).

This document is set up to allow the user of multivariate statistics to get assistance in the choice of multivariate technique to use and the state the data must be in to use the desired technique

2: Summary of Multivariate Statistics

Multivariate statistics provide the ability to analyze complex sets of data. Multivariate statistics provide for analysis where there are many independent (IVs) and possible dependent variables (DVs) which are correlated to each other to varying degrees. The ready availability of software application programs which can handle the complexity of large multivariate data sets has increased and popularized the use of multivariate statistics. The current scientific methodology is increasingly seeking the complex relationships between variables in an attempt to provide for more holistic, inclusive, studies and models. Assessment of results is iterative and stochastic.

For analysis involving multivariate statistics, an appropriate data set is composed of values related to a number of variables for a number of subjects. Accordingly, appropriate data sets may be organized as, a data matrix, a correlation matrix, a variance-covariance matrix, a sum-of-squares and cross-products matrix, or a sequence of residuals (Tabachnick and Fidell, 1989).

3: Data Screening to Assess Validity/Quality of Input

Checklist for screening data. Prior to the fundamental analysis, it is important to consider the data, due to the effect characteristics of the data may have upon the results. Table 1, after Tabachnick and Fidell, (1989), provides for an appropriate sequence for screening the proposed data. The order of the screening is important as decisions at the earlier steps influence decisions to be taken at later steps. For example, if the data is both non-normal and has outliers, the decision to delete values or transform the data is confronted. If transformation is undertaken first, there is likely to be fewer outliers, yet if the outliers are deleted or modified first, there are likely of be fewer variables with non-normality. Transformation of the variable is usually preferred as it typically reduces the number of outliers, and is more likely to produce normality, linearity, and homoscedasticity. Screening of the input data will help assess the appropriateness of the use of a particular data set. Screening will aid in the isolation of data peculiarities and allow the data to be adjusted in advance of further multivariate analysis. The checklist isolates key decision points which need to be assessed to prevent poor data induced analysis problems. Consideration and resolution of problems encountered in the screening of a data set is necessary to ensure a robust statistical assessment.

Table 1. Checklist for Screening Data

1. Inspect univariate descriptive statistics for accuracy of input
 1. out-of-range values, be aware of measurement scales
 2. plausible means and standard deviations
 3. coefficient of variation
2. Evaluate amount and distribution of missing data: deal with problem
3. Independence of variables
4. Identify and deal with nonnormal variables
 1. check skewness and kurtosis, probability plots
 2. transform variables (if desirable)
 3. check results of transformations
5. Identify and deal with outliers
 1. univariate outliers
 2. multivariate outliers
6. Check pairwise plots for nonlinearity and heteroscedasticity
7. Evaluate variables for multicollinearity and singularity

8. Check for spatial autocorrelation

How to deal with the results of screening the data, what to look for.

1. Inspect univariate descriptive statistics for accuracy of input

Generate summary statistics on all the variables and bivariate correlations between all variables. Summary statistics such as mean, median, standard deviation, minimum, maximum. (See transformations page for an example.)

Check range of variables, ensure data values fall within appropriate range, no out-of-range values. In the context of the given data set assess plausible means and standard deviations. Assess the matrix with the coefficient of variation based upon the bivariate correlations, especially if a correlation matrix is to be used as the input data form for multivariate analysis. Correlation's may be inflated, deflated, or inaccurately calculated. Inflated correlations may be due to repetition of a variable, deflated correlations may be due to restricted range of a variable. Observing the bivariate relationships may also foretell variables which are multicollinear or singular, problems that are dealt with in a subsequent section.

1. Evaluate amount and distribution of missing data: deal with problem,
2. Independence of variables, check data for orthogonality
3. Identify and deal with nonnormal variables - normality, skewness, kurtosis. Assess for normality.
4. Identify and deal with outliers.
5. Check pairwise plots for nonlinearity and heteroscedasticity.
6. Evaluate variables for multicollinearity and singularity
7. Check for spatial autocorrelation.

4: Multiple Correlation and Regression

Regression analyses are a set of statistical techniques which allow one to assess the relationship between one dependent variable (DV) and several independent variables (IVs). Multiple regression is an extension of bivariate regression in which several IVs are combined to predict the DV. Regression may be assessed in a variety of manners, such as:

1. *partial regression and correlation*
 - Isolates the specific effect of a particular independent variable controlling for the effects of other independent variables. The relationship between pairs of variables while recognizing the relationship with other variables.
2. *multiple regression and correlation*
 - combined effect of all the variables acting on the dependent variable; for a net,

combined effect. The resulting R² value provides an indication of the goodness of fit of the model. The multivariate regression equation is of the form:

$$Y = A + B_1X_1 + B_2X_2 + \dots + B_kX_k + E$$

where:

Y = the predicted value on the DV,

A = the Y intercept, the value of Y when all Xs are zero,

X = the various IVs,

B = the various coefficients assigned to the IVs during the regression,

E = an error term.

Accordingly, a different Y value is derived for each different case of IV. The goal of the regression is then to derive the B values, the regression coefficients, or beta coefficients. The beta coefficients allow the computation of reasonable Y values with the regression equation, and provide that calculated values are close to actual measured values.

Computation of the regression coefficients provides two major results:

1. minimization of deviations (residuals) between predicted and obtained Y values for the data set,
2. optimization of the correlation between predicted and obtained Y values for the data set.

As a result the correlation between the obtained and predicted values for Y relate the strength of the relationship between the DV and IVs.

Although regression analyses reveal relationships between variables this does not imply that the relationships are causal. Demonstration of causality is not a statistical problem, but an experimental and logical problem.

The ratio of cases to independent variables must be large to avoid a meaningless (perfect) solution. As with more IVs than cases, a regression solution may be found which perfectly predicts the DV for each case. As a rule of thumb there should be approximately 20 times more cases than IVs for good results, yet at a bare minimum 5 times more cases than IVs may be used. Be aware that cases with missing values are generally deleted in the calculation by default. Extreme cases (outliers) have a strong effect on the regression solution and should be dealt with. Calculation of the regression coefficients requires matrix inversion, which is possible only when the variables are not multicollinear or singular. The examination of residual plots will assist in the assessment that the results meet the assumptions of normality, linearity, and homoscedasticity between predicted DV scores and errors of prediction. The assumptions of the analysis are:

- that the residuals (the difference between predicted and obtained scores) are normally distributed,

- that the residuals have a straight line relationship with predicted DV scores, and the variance of the residual about the predicted scores is the same for all predicted scores.

Prior to processing of the data as input to a multiple regression model the data should be screened. The simple mathematics involved and the ubiquity of programs capable of computing regression result in the misuse of regression procedures Wetherill, 1986.

Sample interpretation of beta coefficient table for interpretation of regression results

(Sample table generated with SYSTAT, 1992)

| BETA COEFFICIENT TABLE | | | | | | |
|--------------------------------------|---------------|----------|-------------|---------|--------|-----------|
| DEP VAR: AVGINC | | N: 290 | | | | |
| MULTIPLE R: 0.649 | | | | | | |
| SQUARED MULTIPLE R: 0.421 | | | | | | |
| ADJUSTED SQUARED MULTIPLE R: 0.406 | | | | | | |
| STANDARD ERROR OF ESTIMATE: 5559.626 | | | | | | |
| VARIABLE | COEF. | STD ERR. | STD COEF | TOL. | T | P(2 TAIL) |
| CONSTANT | 47170 | 762.7 | 0.000 | - | 61.85 | 0.000 |
| POP91 | 0.044 | 0.009 | 1.336 | 0.025 | 4.680 | 0.000 |
| NONNAT | -0.020 | 0.026 | -0.195 | 0.031 | -0.761 | 0.447 |
| LOWED | -0.181 | 0.053 | -0.602 | 0.066 | -3.406 | 0.001 |
| HIGHED | -0.079 | 0.112 | -0.321 | 0.010 | -0.707 | 0.480 |
| MUNEMP | -370.2 | 48.16 | -0.358 | 0.948 | -7.688 | 0.000 |
| MINES | -0.011 | 0.181 | -0.004 | 0.612 | -0.061 | 0.951 |
| GOV | 0.092 | 0.115 | 0.119 | 0.092 | 0.799 | 0.425 |
| ANALYSIS OF VARIANCE | | | | | | |
| SOURCE | SUM-OF-SQU'RS | DF | MEAN-SQUARE | F-RATIO | P | |
| REGRES'N | .632959E+10 | 7 | .904228E+09 | 29.254 | 0.000 | |
| RESIDUAL | .871646E+10 | 282 | .309094E+08 | | | |

How to interpret regression results:

- the dependent Y variable should be stated, and the number of variables (ie. AVGINC, N = 290)
- parameters: the intercept, or constant value, (ie. 47170)
- partial regression coefficients, "b" values, (ie. POP91 = 0.044)
- standard error of "b" values, (ie. POP91 = 0.009)
- significance test, such as t-test, of partial regression coefficients, and
- a "p" value, a probability of significance. (significance test and p values are generated for all the beta coefficients and for the model itself)
- significance of model, observe significance test: look for significant F-ratio or t-test, and the corresponding significant p (probability value). Such as, in the ANOVA section of the table generated above
- the F statistic is significant, as is the P value, therefore the model is significant.

- seek which of the independent variables are significant, again look at the t- and P-values, those associated with each independent variable, for significance generally necessary is a high t- value and a low P-value. (Generally a T-value > 2 is significant)
- the beta values, the partial regression coefficients, represent the importance of the variable. As can be seen with MUNEMP, the high beta value, and the highest standard error of beta, also corresponds to significant P- and T-values.
- standard error values - relate the range of the beta values
- the different R-values relate different levels of confidence on the data, the R value is highest, at 0.649, which relates the correlations of the model, while when squared to reflect the amount of variation the IVs and the are explaining in the DV, the value is reduced to 0.421. When controlling for sample size with the adjusted R2 the model accounts for 0.41 of the variation in average family income. The slight change between the R2 and adjusted R2 value reflects the large sample size of the data set. 42% of the variation in average family income, the dependent Y variable, can be explained by the combination of these seven variables.

Some packages, such as SPSS, generate a VIF and tolerance value

- the VIF, or variance inflation factor, will reflect the presence or absence of multicollinearity. A high VIF, larger than one, the variable may be affected by multicollinearity. The VIF has a range 1 to infinity.
- tolerance has a range from zero to one. The closer the tolerance value is to zero relates a level of multicollinearity.

As mentioned above, the results of the regression should be assessed to reflect the quality of the model, especially if the data was not screened. See transformations, linearity, homoscedasticity, residual plot, multivariate normality, and multicollinearity or singularity.

5: Principal Components Analysis and Factor Analysis

Principal components analysis (PCA) and factor analysis (FA) are statistical techniques applied to a single set of variables to discover which sets of variables in the set form coherent subsets that are relatively independent of one another. Variables that are correlated with one another which are also largely independent of other subsets of variables are combined into factors. Factors which are generated are thought to be representative of the underlying processes that have created the correlations among variables.

PCA and FA can be exploratory in nature, FA is used as a tool in attempts to reduce a large set of variables to a more meaningful, smaller set of variables. As both FA and PCA are sensitive to the magnitude of correlations robust comparisons must be made to ensure the quality of the analysis. Accordingly, PCA and FA are sensitive to outliers, missing data, and poor correlations between variables due to poorly distributed variables (See

normality link for more information on distributions.) As a result data transformations have a large impact upon FA and PCA.

Correlation coefficients tend to be less reliable when estimated from small sample sizes. In general it is a minimum to have at least five cases for each observed variable. Missing data need be dealt with to provide for the best possible relationships between variables. Fitting missing data through regression techniques are likely to over fit the data and result in correlations to be unrealistically high and may as a result manufacture factors. Normality provides for an enhanced solution, but some inference may still be derived from nonnormal data. Multivariate normality also implies that the relationships between variables are linear. Linearity is required to ensure that correlation coefficients are generated from appropriate data, meeting the assumptions necessary for the use of the general linear model. Univariate and multivariate outliers need to be screened out due to a heavy influence upon the calculation of correlation coefficients, which in turn has a strong influence on the calculation of factors. In PCA multicollinearity is not a problem as matrix inversion is not required, yet for most forms of FA singularity and multicollinearity is a problem. If the determinant of R and eigenvalues associated with some factors approach zero, multicollinearity or singularity may be present. Deletion of singular or multicollinear variables is required.

Uses of Principle Components Analysis and Factor Analysis

Direct Uses:

- identification of groups of inter-related variables,
- reduction of number of variables,

Indirect Uses:

- a method of transforming data. Transformation of data through rewriting the data with properties the original data did not have. The data may be efficiently simplified prior to a classification while also removing artifacts such as multicollinearity.

Theory to Common Factor Analysis and Factor Analysis

The key underlying base to Common Factor Analysis (PCA and FA) is that the chosen variables can be transformed into linear combinations of an underlying set of hypothesized or unobserved components (factors). Factors may either be associated with 2 or more of the original variables (common factors) or associated with an individual variable (unique factors). Loadings relate the specific association between factors and original variables. Therefore, it is necessary to find the loadings, then solve for the factors, which will approximate the relationship between the original variables and underlying factors. The loadings are derived from the magnitude of eigenvalues associated to individual variables.

The difference between PCA and FA is that for the purposes of matrix computations PCA assumes that all variance is common, with all unique factors set equal to zero; while FA assumes that there is some unique variance. The level of unique variance is dictated by the FA model which is chosen. Accordingly, PCA is a model of a closed system, while FA is a model of an open system.

Rotation attempts to put the factors in a simpler position with respect to the original variables, which aids in the interpretation of factors. Rotation places the factors into positions that only the variables which are distinctly related to a factor will be associated. Varimax, quartimax, and equimax are all orthogonal rotations, while oblique rotations are non-orthogonal. The varimax rotation maximizes the variance of the loadings, and is also the most commonly used.

To run a PCA or FA

To analyze data with either PCA or FA 3 key decisions must be made:

- the factor extraction method,
- the number of factors to extract, and
- the transformation method to be used.

Interpretation of a Factor Analysis

Determination of number of factors to extract

- significance test, difficult to meet assumptions required to significance tests, therefore the following heuristics are used.
- magnitude of eigenvalues,

Assess the amount of original variance accounted for. Retain factors whose eigenvalues are greater than 1. (Ignore those with eigen values less than one as the factor is accounting for less variance than an original variable).

Eigenvalues and Proportion of Original Variance

| | Magnitude | Variance Prop. |
|----------|-----------|----------------|
| Value 1 | 11.181 | .532 |
| Value 2 | 3.141 | .15 |
| Value 3 | 1.712 | .082 |
| Value 4 | 1.255 | .06 |
| Value 5 | 1.074 | .051 |
| Value 6 | .78 | .037 |
| Value 7 | .478 | .023 |
| Value 8 | .394 | .019 |
| Value 9 | .363 | .017 |
| Value 10 | .217 | .01 |
| Value 11 | .183 | .009 |

the figure relates that five factors are significant.

- substantive importance,
an absolute test of eigen values in a proportional sense. Retain any eigenvalue that accounts for at least 5% of the variance,
- skree test,
plot magnitude of eigen values (Y axis) versus components (X axis), retain factors which are above the inflection point of the slope.
- interpretability,
a battery of tests where the above heuristics may all be applied, assess magnitude of eigen values, substantive importance, and a skree test.

Which variables are best summarized by the model?

- interpret communalities (final estimates of communalities),
high = most important,
low = least important.

Communality Summary

| | SMC | Final Estimate | | SMC | Final Estimate |
|---------|------|----------------|--------|------|----------------|
| POP86 | 1 | .995 | HOUSEV | .806 | .866 |
| POP91 | 1 | .993 | CROWD | .723 | .813 |
| POPDEN | .923 | .902 | PPFAM | .635 | .784 |
| CDAREA | .498 | .775 | AVGINC | .725 | .798 |
| FRENCH | .943 | .727 | LOWINC | .997 | .978 |
| NONNAT | .991 | .881 | | | |
| NONMOV1 | 1 | .993 | | | |
| NONMOV5 | .999 | .989 | | | |
| LOWED | .996 | .966 | | | |
| HIGHED | .993 | .974 | | | |
| MUNEMP | .895 | .94 | | | |
| FUNEMP | .884 | .929 | | | |
| LABOUR | 1 | .991 | | | |
| AGRIC | .671 | .746 | | | |
| MINES | .614 | .609 | | | |
| GOV | .938 | .716 | | | |

The communalities relate the overall effect of the factors.

Naming of Factors

- look at individual factor scores, see which variables have the highest factor scores. Also look at the factor scores to see if the initial interpretations are confirmed by the factor scores (Factor scores are normally distributed only when the input variables are normally distributed. Therefore, when interpreting factors the greatest concern is with the tail values. The normal distribution of factor scores also acts as a data transformation and prepares the data for other multivariate analyses.)

What is meant by an Ill conditioned Correlation Matrix

- an illconditioned correlation matrix is a manifestation of multicollinearity. FA is sensitive to an illconditioned matrix while PCA is not. To solve for the characteristic equation in FA matrix inversion is required, which is not possible with a singular matrix. To solve this problem click on the multicollinearity link.

To assess the value of input variables to the model

- assess the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO), which provides results in the range from 0.5 to 0.9.

Measures of Variable Sampling Adequacy
Total matrix sampling adequacy : .833

| | | | |
|---------|------|--------|------|
| POP86 | .865 | HOUSEV | .884 |
| POP91 | .824 | CROWD | .647 |
| POPDEN | .911 | PPFAM | .628 |
| CDAREA | .629 | AVGINC | .821 |
| FRENCH | .693 | LOWINC | .88 |
| NONNAT | .81 | | |
| NONMOV1 | .82 | | |
| NONMOV5 | .852 | | |
| LOWED | .848 | | |
| HIGHED | .901 | | |
| MUNEMP | .645 | | |
| FUNEMP | .674 | | |
| LABOUR | .911 | | |
| AGRIC | .758 | | |
| MINES | .534 | | |
| GOV | .85 | | |

Bartlett Test of Sphericity- DF: 230 Chi Square: 16626.24 P: ●

For example: MINES is a less valid variable than POPDEN in this model.

A value of 1 relates a complete relationship, totally related, which is bad. The range which is provided as a heuristic is:

- 0.9 - marvelous,
- 0.8 - meritorious,
- 0.7 - middling,
- 0.6 - mediocre, or
- 0.5 - miserable (perfectly uncorrelated).

Barlett test of sphericity, variable projected upon an n-dimensional spheroid, the significance of the relationship is then evaluated. (See figure above, where the p value is significant, and does not fit in the allocated space).

6: Discriminant Analysis

The main use of discriminant analysis is to predict group membership from a set of predictors. Discriminant function analysis consists of finding a transform which gives the maximum ratio of difference between a pair of group multivariate means to the multivariate variance within the two groups (Davis, 1986). Accordingly, an attempt is made to delineate based upon maximizing between group variance while minimizing within group variance. The predictors characteristics are related to form groups based upon similarities of distribution n-dimensional space which are then compared to groups which are input by the user as truth. This enables the user to test the validity of groups based upon actual data, to test groups which have been created, or to put objects into groups. Discriminant analysis (DA) may act as a univariate regression and is also related to ANOVA (Wesolowsky, 1976). The relationship to ANOVA is such that DA may be considered as a multivariate version of ANOVA. The underlying assumptions of DA are:

- the observations are a random sample,
- each group is normally distributed, DA is relatively robust to departures from normality,
- the variance/covariance matrix for each group is the same,
- each of the observations in the initial classification is correctly classified (training data).

Prior probabilities may be assigned to deal with unequal sample sizes. Although the size of the smallest group should still be larger than the number of predictor variables, at a minimum. The assumption of multivariate normality holds, that the scores on predictors are randomly distributed, and that the sampling distribution is of any linear combination of predictors is linearly distributed. Discriminant analysis is relatively robust to nonnormality due to skewness, yet not that which is due to outliers. Discriminant analysis is highly sensitive to outliers. Variables with significant outliers necessitate transformation prior to analysis. Linearity is also assumed for discriminant analysis. The inclusion of redundant variables in the computation of discriminant analysis normally results in automatic exclusion of multicollinear and singular variables. A tolerance test is undertaken to assess the viability of all independent variables prior to analysis with most statistical application programs.

Analysis of Output:

Importance of variables to model:

- test of significance: D^2 , if the discriminant functions discriminate well, the D^2 will be large, yet if the D^2 is small, the discriminant functions do not discriminate well.
- in the context of a oneway ANOVA, the Wilks' Lambda is a reflectance of a variables importance. The smaller the Wilks' Lambda, the more important the variable. An F-test and significant p value are also provide as measures of the importance of a variable. Therefore, the smallest Wilks' Lambda and the greatest

significance relates the most important variable. This relationship is demonstrated in the following graphic, where CDAREA has the lowest Wilks' Lambda, the highest F-ratio, and a significant p value reflecting the importance of this variable to the analysis.

Wilks' Lambda (U-statistic) and univariate F-ratio with 11 and 278 degrees of freedom

| Variable | Wilks' Lambda | F | Significance |
|----------|---------------|---------|--------------|
| LABOUR | .93557 | 1.7984 | .0540 |
| LOWED | .96595 | .8909 | .5497 |
| LOWINC | .97622 | .6156 | .8152 |
| MINES | .82848 | 5.2320 | .0000 |
| MUNEMP | .50920 | 24.3592 | .0000 |
| NONMOV1 | .95237 | 1.8331 | .0485 |
| NONMOVS | .93760 | 1.6318 | .0771 |
| NONNAT | .94604 | 1.4414 | .1539 |
| POP86 | .94204 | 1.5551 | .1119 |
| POP91 | .93470 | 1.7655 | .0598 |
| POP9EN | .96975 | .7885 | .6516 |
| PPFAM | .62109 | 15.4183 | .0000 |
| AGRIC | .52938 | 22.4675 | .0000 |
| AVGINC | .66401 | 12.7882 | .0009 |
| CDAREA | .34655 | 47.6548 | .0000 |
| CROWD | .61153 | 16.0545 | .0000 |
| FRENCH | .87338 | 3.6640 | .0001 |
| FUNEMP | .38854 | 39.7719 | .0000 |
| SOV | .94374 | 1.5066 | .1284 |
| HIGHED | .95461 | 1.2016 | .2855 |
| HOUSEV | .49201 | 26.0939 | .0000 |

SPSS Output, Key Features

- the groups are defined by the number of cases in each group, and the new weights associated with the variable to create or account for biases due to sample size or research aims,
- Wilks' Lambda, F-ratio, and significance for each of the variables,
- variables which failed the tolerance test will be presented,
- the Canonical Discriminant Functions Table: relates the number of important functions through a variety of tests:
 - eigenvalues, which are greater 1 are significant,
 - percentage of variance greater than 5 is significant,
 - Cumulative percentage to approximately 75% is significant,
 - a canonical correlation of greater than 0.6 is significant, and
 - a probability p value.

The preceding criteria are all presented on the following graphic. What may be seen is that each of the tests have different cut-off points for the number of significant functions, accordingly use discretion and apply a value which is appropriate given the results and nature of input data.

| Canonical Discriminant Functions | | | | | | | | | |
|----------------------------------|------------|-----------------|---------|----------------|-----------|---------------|------------|-----|-------|
| Fcn | Eigenvalue | Pct of Variance | Cum Pct | Canonical Corr | After Fcn | Wilks' Lambda | Chi-square | df | Sig |
| 1* | 3.0021 | 34.73 | 34.73 | .8661 | 0 | .006516 | 1379.188 | 198 | .0000 |
| 2* | 2.0417 | 23.62 | 58.35 | .8193 | 1 | .026077 | 999.199 | 170 | .0000 |
| 3* | 1.6158 | 18.69 | 77.04 | .7659 | 2 | .079317 | 694.401 | 144 | .0000 |
| 4* | .9280 | 10.74 | 87.78 | .6938 | 3 | .207459 | 430.953 | 120 | .0000 |
| 5* | .4670 | 5.40 | 93.18 | .5642 | 4 | .399990 | 251.070 | 96 | .0000 |
| 6* | .3361 | 3.89 | 97.07 | .5010 | 5 | .586782 | 146.070 | 78 | .0000 |
| 7* | .1234 | 1.43 | 98.50 | .3314 | 6 | .784024 | 66.869 | 60 | .2585 |
| 8* | .0564 | .65 | 99.15 | .2311 | 7 | .880766 | 34.788 | 44 | .8384 |
| 9* | .0469 | .54 | 99.69 | .2117 | 8 | .930444 | 19.754 | 30 | .9228 |
| 10* | .0206 | .24 | 99.93 | .1420 | 9 | .974085 | 7.194 | 18 | .9884 |
| 11* | .0059 | .07 | 100.00 | .0767 | 10 | .994119 | 1.616 | 8 | .9906 |

- structure matrix, which presents the discriminant functions related to each variable. The largest absolute correlation between each variable and discriminant functions are denoted and this assists in the naming of significant functions,
- a territorial map is also presented which relates the delineation of groups around centroids, the quality of the groups is presented in this 2-Dimensional display,
- list of cases, with membership probabilities. For example:

| Case # | Group | Actual Group | Highest P(D G) | Prob P(G D) | 2nd High Group | P(G D) | Discrm. Scores |
|--------|-------|--------------|----------------|-------------|----------------|--------|----------------|
| 1 | 1 | 1 | 0.0002 | 0.8677 | 3 | 0.6968 | -1.685 |
| 6 | 1** | 5 | 0.6828 | 0.6819 | 4 | 0.1873 | -1.008 |

Case 1 is a case which is moderately specified on the periphery of group 1, by the training data, with the actual case being the same as that dictated by the highest probability. The P(D|G), the probability of the dictated class (actual class) given the probability of the calculated group, the probability of D given G (a measure of how central an object is to the group, the conditional probability). Yet for case number 6, the group with the highest probability of being in the correct class based upon the input data is not the same as the class provided as truth. P(G|D) is the posterior probability is, given that a discriminant score D, what is the probability the object is in G. The actual class was not even second highest group. This table enables users to assess the validity of a grouping structure and to also screen for problem cases in an analysis. Problem cases which have similar characteristics may require a new variable to assist in the delineation of the groups.

Classification results - a table of actual group (Y) versus predicted group membership (X) which tallies for the total number of cases where they were grouped. Ideally, with a 100% accurate classification the trace, down the matrix diagonal, would be all ones. Yet, with a non-perfect matrix, the class where the values are being placed is visible. A tally of the results of the previous case placement table in a percentage form.

As can be expected there is more than one way to calculate the accuracy of the classification. Congalton, (1991), presents two methods of calculation classification

accuracy: users accuracy and producers accuracy. Users accuracy calculates correctly classed from the trace variable over the row total and provides an indication of errors of case omission. Producers accuracy is the calculation of correctly classed from the trace value over the column total. Producers accuracy gives an indication of the accuracy of what the model was able to itself predict, whereas users accuracy relates how well the training data was discerned. The goal of the analysis ought to dictate which of the methods is utilized.

| | NFLD | PEI | NS | NB | PQ | ON | MN | SK | AB | BC | YT | NWT | TOTAL |
|-------|------|-----|----|----|-----|----|----|----|----|----|----|-----|-------|
| NFLD | 9 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| PEI | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |
| NS | 1 | 0 | 9 | 0 | 7 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 18 |
| NB | 0 | 0 | 5 | 2 | 7 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 15 |
| PQ | 2 | 0 | 1 | 0 | 85 | 1 | 6 | 0 | 0 | 3 | 0 | 1 | 99 |
| ON | 0 | 0 | 0 | 0 | 4 | 44 | 0 | 0 | 0 | 0 | 1 | 0 | 49 |
| MN | 0 | 0 | 0 | 0 | 2 | 0 | 19 | 1 | 0 | 1 | 0 | 0 | 23 |
| SK | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 15 | 0 | 0 | 0 | 0 | 18 |
| AB | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 5 | 9 | 1 | 0 | 0 | 18 |
| BC | 0 | 0 | 0 | 0 | 7 | 3 | 1 | 0 | 1 | 18 | 0 | 0 | 30 |
| YT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| NWT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 3 | 5 |
| TOTAL | 12 | 1 | 16 | 2 | 115 | 48 | 33 | 22 | 10 | 23 | 4 | 4 | 290 |

The diagonal of the confusion matrix is the number of variables which were both trained and predicted to the same group. Division of the percentage accuracy may be calculated either with the row total or column total. The row total is the number of variables trained to be in the class. The column total is the number of variables *predicted* to be in that class. Accordingly, the different results both have uses valid to different situations. The row division method is referred to as USERS accuracy, and relates the amount of commission from the class. The column division method is called producers accuracy, and relates the amount of omission error. The purpose of an individual study will dictate which manner of accuracy should be applied. (Commission error = consumers accuracy; omission error = producers accuracy)

The preceding graphic demonstrates the profoundly different accuracy results which will be calculated based upon use of either users or producers accuracy. Of significant note are the cases or groups with small sample sizes.

7: Cluster Analysis

Cluster analysis (CA) is a multivariate procedure for detecting natural groupings in data. Cluster analysis classification is based upon the placing of objects into more or less homogeneous groups, in a manner such that the relationship between groups is revealed. CA lacks an underlying body of statistical theory and is heuristic in nature. Cluster analysis requires decisions to be made by the user relating to the calculation of clusters, decisions which have a strong influence on the results of the classification. CA is useful to classify groups or objects and is more objective than subjective.

Clustering methods may be top down and employ logical division, or bottom up and undertake aggregation. Aggregation procedures which are based upon combining cases

through assessment of similarities are the most common and popular will be the focus of this section.

Care should be taken that groups (classes) are meaningful in some fashion and are not arbitrary or artificial. To do so the clustering techniques attempt to have more in common with own group than with other groups, through minimization of internal variation while maximizing variation between groups. Homogeneous and distinct groups are delineated based upon assessment of distances or in the case of Ward's method, an F-test (Davis, 1986).

Steps to Cluster Analysis

The two key steps within cluster analysis are the measurement of distances between objects and to group the objects based upon the resultant distances (linkages). The distances provide for a measure of similarity between objects and may be measured in a variety of ways, such as Euclidean and Manhattan metric distance. The criteria used to then link (group) the variables may also be undertaken in a variety of manners, as a result significant variation in results may be seen. Linkages are based upon how the association between groups is measured. For example, simple linkage or nearest neighbor distance, measures the distance to the nearest object in a group while furthest neighbor linkage or complete linkage, measures the distance between furthest objects. These linkages are both based upon single data values within groups, whereas average between group linkage is based upon the distance from *all* objects in a group. Centroid linkage has a new value, representing the group centroid, which is compared to the ungrouped point to weigh inclusion. Ward's method is variance based with the groups variance assessed to enable clustering. The group which see the smallest increase in variance with the iterative inclusion of a case will receive the case. Ward's is a popular default linkage which produces compact groups of well distributed size. Standardization of variables is undertaken to enable the comparison of variables to minimize the bias in weighting which may result from differing measurement scales and ranges. Z score format accounts for differences between mean values and reduces the standard deviation when variables have multivariate normality. Multicollinearity will bias the clusters due to the high correlations between variables.

Choosing number of groups

The ideal number of groups to establish may be assessed graphically or numerically. Graphically the number of groups may be assessed with an icicle plot or dendrogram. The dendrogram bisected at a point which will divide the cases into a cluster based upon groupings up to the point where the bisection occurred. Numerically the number of cases may be assessed on the agglomeration schedule, by counting up from the bottom to where a significant break in slope (numbers) occurs. This is similar to a visual interpretation of a skree plot

The optimal number of groups may be assessed *a priori* based upon knowledge of the data set. A skree plot which converts a dendrogram to a profile curve will have an

extreme inflection point where the number of groups significantly changes. The number of groups above the inflection point is an appropriate number of groups. Optimality of classes may be assessed by how "natural" the classes appear. Low within class variation in comparison to the between class variation reflects an appropriate class structure. Discriminant analysis may also be employed to assess optimality and efficiency of computed groups, by imputing the cluster analysis derived classes for analysis with the original data.

8: Spatial Autocorrelation

Spatial autocorrelation is an assessment of the correlation of a variable in reference to spatial location of the variable. Assess if the values are interrelated, and if so is there a spatial pattern to the correlation, there is spatial autocorrelation. Spatial autocorrelation measures the level of interdependence between the variables, the nature and strength of the interdependence. Spatial autocorrelation may be classified as either positive or negative. Positive spatial autocorrelation has all similar values appearing together, while negative spatial autocorrelation has dissimilar values appearing in close association.

Spatial autocorrelation is related to the scale of the data as a periodicity of elements is assessed. Negative spatial autocorrelation is more sensitive to changes in scale. In geographic applications there is usually positive spatial autocorrelation.

Uses of assessment of spatial autocorrelation:

- identification of patterns which may reveal an underlying process,
- describe a spatial pattern and use as evidence, such as a diagnostic tool for the nature of residuals in a regression analysis,
- as an inferential statistic to buttress assumptions about the data,
- data interpolation technique.

How to Compute Spatial Autocorrelation:

- The measurement scale dictates the type of measure,
- assign weights to the cases,
- create a matrix representing the relationships between variables into software such as Anaspace. Anaspace, by Michael Tiefelsdorf, of Wilfrid Laurier University, which will compute measure of the spatial autocorrelation between the input data matrix. In the case of raster data, when building a contiguity matrix be cognizant of the relationships between cases, are neighbors determined on a eight directional Queens case or non-diagonal, four directional Rooks case.

How to Measure Spatial Autocorrelation:

Moran's I

Computation of Moran's I is achieved by division of the spatial covariation by the total variation. Resultant values are in the range from approximately -1 to 1. Positive signage represents positive spatial autocorrelation, while the converse is true for negative signage. With a Zero result representing no spatial autocorrelation.

Geary's C

Computation of Geary's C results in a value within the range of 0 to +2. With zero being a strong positive spatial autocorrelation, through to 2 which represents a strong negative spatial autocorrelation.

How to Correct for Spatial Autocorrelation in Regression:

- indicates incomplete model, there may be a missing variable. Therefore add an additional variable which may change data pattern.
- incorrect model specification. The data may not be appropriate for a linear fit, or a non-spatial effect may be manifest in the residuals, nuisance spatial autocorrelation. Substantive spatial autocorrelation occurs when there is missing values.
- dominant or extreme cases, outliers, which should have been found at data screening stage.
- systematic measurement error in response variable (non-random). A case in which error increases as values increase, or *vice versa*.
- regression model is inappropriate, reflects the need for an explicitly spatial model. A spatially autoregressive model which incorporates a spatial lag operator into the regression computation. The approach for the implementation of spatial autoregressive models is as follows:
 - establish nature of spatial dependency,
 - use information to choose appropriate model form,
 - fit model using maximum likelihood operators,
 - calculate residuals from model,
 - test residuals, and
 - adjust model based upon residuals. (after Haining, 1990)

Appendixes

1. Census Data

D_R CODE CDNAME POP86 POP91 POPDEN CDAREA FRENCH NONNAT
NONMOV1 NONMOV5 LOWED HIGHED MUNEMP FUNEMP LABOUR AGRIC
MINES GOV HOUSEV CROWD PPFAM AVGINC LOWINC ID PROV

1001000 DIVISION NO. 1 246149 253203 27.65 9158.8 560 5350 213705 153940
28440 18155 22.8 20.6 125035 1035 395 16845 83052 0.3 3.2 45516 10325 1 1

1002000 DIVISION NO. 2 30285 29345 4.53 6477.8 55 210 26715 22340 6105 860 28.7
32.5 12910 60 115 1115 47798 0.5 3.5 35500 1225 2 1

1003000 DIVISION NO. 3 25737 24236 1.14 21259.1 15 85 22415 19285 6725 635 30
37.5 10980 0 55 1075 46274 0.5 3.4 33530 1065 3 1

1004000 DIVISION NO. 4 27278 25691 3.37 7624.6 695 285 21820 17005 4685 880
38.7 35.7 10915 145 65 1590 49344 0.5 3.2 32533 1855 4 1

1005000 DIVISION NO. 5 45648 45314 4.84 9355.6 90 630 38985 29765 6295 2155
27.2 27 21295 340 150 1925 69676 0.5 3.2 40767 1855 5 1

1006000 DIVISION NO. 6 40714 40236 2.25 17895.9 200 640 33955 24195 5100 1820
24.6 20.7 18975 180 85 2580 67761 0.5 3.2 42530 1530 6 1

1007000 DIVISION NO. 7 43618 43170 4.3 10039.8 40 240 38855 32430 9575 1200
40.5 40.9 18690 175 65 1450 45047 0.5 3.2 33055 1860 7 1

1008000 DIVISION NO. 8 54225 51882 5.31 9776 45 250 47385 40090 12425 1300
46.9 40.7 21085 165 670 1615 40886 0.5 3.4 32375 2410 8 1

1009000 DIVISION NO. 9 25954 25022 1.71 14609.1 20 155 22880 19465 6210 685 38
40.2 12110 25 140 965 41057 0.5 3.4 36146 895 9 1

1010000 DIVISION NO. 10 28741 30375 0.11 265437.4 650 620 25165 17375 3575
1200 19.7 26.8 15155 25 2750 2680 46385 0.5 3.5 50854 805 10 1

1101000 KINGS COUNTY 19509 19328 11.54 1675.2 175 525 17455 13285 2790 720
13.1 12.3 9935 900 20 865 65075 0.4 3.2 41719 490 11 2

1102000 QUEENS COUNTY 63460 67196 33.48 2007.2 1205 2765 55230 36695 5395
5940 10.6 12.7 35845 2405 25 4920 90183 0.3 3.2 46059 1965 12 2

1103000 PRINCE COUNTY 43677 43241 21.86 1977.8 4075 815 37210 26935 7070
1715 14.8 18.7 22500 2230 165 2325 59669 0.4 3.2 39811 1105 13 2

1201000 SHELBURNE COUNTY 17516 17343 7.36 2356.5 210 470 15450 11895 3075
610 11.1 18.2 8370 55 10 855 61916 0.4 3 41593 555 14 3

1202000 YARMOUTH COUNTY 27073 27891 13.47 2070.7 6300 960 23640 17040
4830 1120 11 15.3 13025 175 170 825 69627 0.4 3 39546 1080 15 3

1203000 DIGBY COUNTY 21852 21250 8.59 2472.4 6700 570 18800 15015 3900 885
15.3 20 10215 345 0 780 58230 0.4 3 37386 760 16 3

1204000 QUEENS COUNTY 13125 12923 5.46 2367.7 70 475 11515 8775 2355 555
12.5 12.8 5710 55 0 390 59979 0.4 3 38641 515 17 3

1205000 ANNAPOLIS COUNTY 23589 23641 7.38 3203.6 355 1070 19705 14085
2545 1200 10.5 13.6 11285 555 30 2265 66064 0.4 3 36407 935 18 3

1206000 LUNENBURG COUNTY 46483 47634 16.54 2880.4 385 1820 41575 30005
7750 2395 10.1 11.1 22390 605 30 1460 80240 0.4 2.9 39421 1590 19 3

1207000 KINGS COUNTY 53275 56317 25.81 2182.2 1135 2545 45680 28865 5285
4440 9.3 10.8 28240 2400 75 3660 82924 0.4 3 40706 1770 20 3

1208000 HANTS COUNTY 36548 37843 12.39 3054.7 350 1425 32785 22985 4200
1860 13.1 11.5 18615 1130 400 1605 73567 0.4 3 41885 1150 21 3

1209000 HALIFAX COUNTY 306418 330846 59.53 5557.2 8740 20980 258015
152445 21140 43120 8.8 10 184665 830 645 30125 109383 0.3 3 51370 10510 22 3

1210000 COLCHESTER COUNTY 45093 47683 13.16 3622.3 415 1745 39785 26040
4140 3080 13.6 14.1 23550 965 220 1975 71452 0.4 3 40620 1660 23 3

1211000 CUMBERLAND COUNTY 34819 34284 8 4288.1 360 1075 28990 21460
3985 1390 16.5 16.7 15555 760 395 1485 54541 0.4 3 37153 1330 24 3

1212000 PICTOU COUNTY 49772 49651 17.9 2774.4 500 1535 42815 30995 5305
2560 14.8 15 23160 610 135 1220 62144 0.4 3.2 42387 1785 25 3

1213000 GUYSBOROUGH COUNTY 12568 11724 2.68 4370.8 195 185 10845 9050
2490 335 19.1 17.3 4995 85 10 425 43033 0.5 3.2 35186 430 26 3

1214000 ANTIGONISH COUNTY 18929 19226 13.07 1471.2 715 630 16770 12565
1435 1750 14 13.6 8925 365 45 480 69299 0.5 3.5 45645 485 27 3

1215000 INVERNESS COUNTY 21946 21620 5.85 3693.7 3230 595 18940 15150 2530
1225 18 20.3 10090 235 220 635 57196 0.5 3.4 41488 630 28 3

1216000 RICHMOND COUNTY 11841 11260 9.13 1233.4 3175 235 10240 8460 1940
430 20.1 23.6 4715 75 35 365 48766 0.5 3.2 35286 350 29 3

1217000 CAPE BRETON COUNTY 123625 120098 48.57 2472.9 1100 2515 104925
79315 14995 6135 19.8 19.3 49790 285 2660 4335 57848 0.5 3.1 38906 6055 30 3

1218000 VICTORIA COUNTY 8704 8708 3.15 2767.8 65 270 7965 6525 1330 380 19.7
27 4225 70 60 410 59220 0.4 3.2 40476 265 31 3

1301000 SAINT JOHN COUNTY 82460 81462 52.27 1558.6 3910 3195 64430 40615
9245 4660 12.1 12.3 39825 140 60 2960 73842 0.4 3 42502 3930 32 4

1302000 CHARLOTTE COUNTY 26525 26607 7.92 3360.4 620 1455 23500 17720
3955 1350 15.6 18.6 12875 200 15 925 61720 0.4 3 39496 875 33 4

1303000 SUNBURY COUNTY 22894 23575 8.52 2767.9 1710 980 18305 10715 2435
1080 8.1 15.8 12725 240 90 4720 64310 0.5 3 41711 470 34 4

1304000 QUEENS COUNTY 12487 12519 3.4 3680 640 350 10965 8405 2360 380 19.3
15.8 5510 265 255 555 51342 0.4 3 35566 590 35 4

1305000 KINGS COUNTY 56598 62122 17.43 3563.9 1875 2935 53745 36110 5280
4690 9.1 11.3 30855 1000 1045 2035 82576 0.4 3.2 50110 1680 36 4

1306000 ALBERT COUNTY 24832 25640 14.32 1789.9 1265 875 22095 14530 1955
1745 10.6 12.8 13660 220 30 1180 74996 0.4 3.2 48738 610 37 4

1307000 WESTMORLAND COUNTY 110969 114745 30.86 3718.1 45855 3630 94870
64690 16240 8980 12.2 12 58775 855 145 5615 75289 0.3 3 44260 4220 38 4

1308000 KENT COUNTY 31496 31694 7.02 4516.2 23895 1140 28380 22375 8335 835
27.8 21 14985 335 175 1270 52252 0.5 3.2 35380 1185 39 4

1309000 NORTHUMBERLAND COUNTY 52981 52983 4.33 12228.4 14170 840
46460 35735 9350 2375 25.2 22.6 24205 245 425 2785 61663 0.5 3.2 39224 2190 40 4

1310000 YORK COUNTY 77211 82326 9.03 9120.8 4790 4680 67970 45235 7755
11155 11.2 12 44985 915 65 7295 88647 0.3 3 49473 2505 41 4

1311000 CARLETON COUNTY 25429 26026 7.97 3267.3 350 1035 22760 16640 4105
1150 10.5 11.8 12590 1410 10 710 60602 0.4 3.2 39554 995 42 4

1312000 VICTORIA COUNTY 21504 20786 3.8 5474.6 8755 695 18425 14085 3850
1015 14 19.2 9335 795 20 590 61093 0.4 3.2 35256 970 43 4

1313000 MADAWASKA COUNTY 36662 36554 10.68 3421.8 33755 1085 31770
23540 7445 2145 18.1 16.5 16650 475 0 1115 62225 0.5 3 38032 1605 44 4

1314000 RESTIGOUCHE COUNTY 39921 38760 4.6 8430.7 23315 345 33405 25055
8030 1670 22.2 20.2 17540 290 125 1305 59029 0.5 3 38300 1780 45 4

1315000 GLOUCESTER COUNTY 87473 88101 18.87 4669.8 71455 735 78010 60745
19915 4120 23.2 20.2 41170 480 2070 2765 59473 0.5 3.1 36455 4415 46 4

2401000 LES ILES-DE-LA-MADELEINE 14532 13991 69.16 202.3 13060 35 12575
10505 3670 555 19.5 25.7 7540 65 230 430 53141 0.5 3 41297 455 47 5

2402000 PABOK 23758 21713 7.06 3075.3 19585 50 19625 15930 6120 570 30.2 30.1
9725 75 30 780 51416 0.5 3 35410 1430 48 5

2403000 LA COTE-DE-GASPE 22833 20903 5.05 4143.2 17950 75 18550 14185 4740
885 23.3 20.5 9545 105 370 740 55662 0.5 3.2 38106 945 49 5

2404000 DENIS-RIVERIN 15241 14019 2.72 5156.2 13770 30 12200 9525 4015 340
36.2 29.7 5910 125 180 600 47077 0.4 3 31437 1020 50 5

2405000 BONAVENTURE 20616 19848 4.52 4392.5 16455 80 17725 13900 4365 720
22 20.2 8885 270 35 665 55579 0.4 3 38429 755 51 5

2406000 AVIGNON 15475 15494 4.25 3645.4 12650 135 13205 9940 3250 620 23.3
18.8 6525 110 10 380 55057 0.5 3 37198 500 52 5

2407000 LA MATAPEDIA 22055 20930 3.88 5387.8 20775 60 18060 13315 4925 600
28 23.7 8850 740 10 395 45429 0.5 3.2 34559 935 53 5

2408000 MATANE 25258 24334 7.35 3311.6 24075 65 21040 15530 5415 925 17.7
15.6 10955 425 45 830 54283 0.4 3 38263 1230 54 5

2409000 LA MITIS 21762 20157 8.79 2293.6 19750 50 17185 12620 4130 740 16.6 14
8780 725 30 590 53753 0.4 3 36864 925 55 5

2410000 RIMOUSKI-NEIGETTE 50108 51290 19.92 2575.1 50610 320 43005 28320
7230 4360 12.6 12.6 26805 860 120 2210 70221 0.4 3 45307 1865 56 5

2411000 LES BASQUES 11320 10325 8.38 1231.7 10280 30 9085 7125 2780 360 19.1
16.7 4310 450 15 210 41889 0.4 3 32638 590 57 5

2412000 RIVIERE-DU-LOUP 31001 31485 24.81 1269.3 31190 140 27075 18615 5375
1500 12.3 13.1 14805 815 320 765 68780 0.4 3 41344 1155 58 5

2413000 TEMISCOUATA 24795 23348 6.03 3874.3 23115 265 20855 15900 5660 725
20.3 19 9505 495 30 500 47113 0.5 3.2 33442 1355 59 5

2414000 KAMOURASKA 24535 23268 12.06 1929.3 23060 145 20660 15660 4920
1010 17.6 14.8 10560 1385 165 475 59726 0.4 3.2 37320 990 60 5

2415000 CHARLEVOIX-EST 18177 17413 7.33 2375.1 17275 65 15455 11755 3680
475 16.3 17.7 7595 480 0 485 73481 0.5 3 36847 830 61 5

2416000 CHARLEVOIX 13843 13547 3.61 3756.6 13440 25 11885 9425 3160 435 15.1
17.2 6000 325 45 265 65874 0.5 3 37432 460 62 5

2417000 L'ISLET 21189 19938 9.53 2091.6 19795 50 18145 14430 5500 625 10.6 10.1
8840 840 10 390 52147 0.4 3.2 34472 755 63 5

2418000 MONTMAGNY 24794 23667 14.1 1678.2 23370 95 21050 16075 6240 825
11.6 12.3 10980 695 40 600 60641 0.4 3 37467 1090 64 5

2419000 BELLECHASSE 29932 29475 18.09 1629.4 29035 195 26185 20455 7085
1060 7.5 10.3 13865 2035 45 500 64212 0.5 3.2 41479 770 65 5

2420000 L'ILE-D'ORLEANS 6769 6938 35.6 194.9 6810 125 6345 4900 1000 495 6
10.5 3630 495 0 445 111725 0.3 3 47653 160 66 5

2421000 LA COTE-DE-BEAUPRE 20563 21214 4.26 4985.1 20900 70 18665 14115
4025 1020 9.1 11 10820 220 35 1185 78952 0.5 3 45438 725 67 5

2422000 LA JAKUES-CARTIER 20467 23282 7.32 3179.9 21220 500 18430 11070
2245 1655 7 12.6 12645 130 15 3625 91878 0.5 3 49232 490 68 5

2423000 COMMUNAUTE URBAINE DE QUEBEC 464865 490271 902.23 543.4
468535 12135 397710 246655 59065 58700 9.2 9.6 265055 1510 350 43770 100076 0.5
2.8 49271 22975 69 5

2424000 DESJARDINS 46398 49076 193.44 253.7 48130 390 41310 27660 7085 3570
7 10 25605 550 50 1845 84530 0.5 3 46349 1980 70 5

2425000 LES CHUTES-DE-LA-CHAUDIERE 56920 67479 161.2 418.6 65930 840
57475 35160 5575 6555 6.3 8.3 36585 795 55 3915 89085 0.5 3.2 50750 1965 71 5

2426000 LA NOUVELLE-BEAUCE 23164 24362 30.73 792.8 24050 150 21660 16550
4545 850 6.1 9.6 12210 1765 15 390 72189 0.5 3.2 42099 630 72 5

2427000 ROBERT-CLICHE 18717 18586 22.69 819 18420 70 16800 13050 4080 520
8.6 9.1 8630 1110 30 345 58836 0.5 3.2 39088 635 73 5

2428000 LES ETCHEMINS 19485 18668 10.34 1805.5 18490 120 16780 13260 5340
415 18.2 13.6 7920 395 15 340 48590 0.5 3.2 34642 850 74 5

2429000 BEAUCE-SARTIGAN 41655 44218 22.8 1939.3 43675 505 38295 27700 8985
1700 12 9.6 21685 1030 45 685 64822 0.5 3.2 40114 1500 75 5

2430000 LE GRANIT 20782 20993 7.68 2733.3 20495 300 18735 13925 5375 545 10
9.3 10170 850 65 385 59551 0.5 3.2 35960 815 76 5

2431000 L'AMIANTE 48327 45851 24.07 1904.7 44805 550 40595 31195 9635 1585
9.3 9.6 20705 1235 1675 730 58925 0.5 3 39401 1755 77 5

2432000 L'ERABLE 25389 24680 19.12 1291 24270 255 21185 15355 5275 760 10.5
8.8 11855 1320 20 340 60841 0.5 3.2 38362 715 78 5

2433000 LOTBINIERE 26187 26633 16.17 1647.2 26135 200 23900 17630 5805 955
9.8 7.5 12975 2170 0 640 61888 0.5 3.2 39089 770 79 5

2434000 PORTNEUF 41622 43179 11 3924.3 42460 205 37800 28470 9495 1705 11
11.8 20090 1445 170 1790 66383 0.4 3 39602 1495 80 5

2435000 MEKINAC 13899 13629 2.46 5544.3 13480 110 12355 9605 3415 350 21.6
16.7 6025 500 25 300 49322 0.4 3 36323 595 81 5

2436000 LE CENTRE-DE-LA-MAURICIE 67089 67379 52.18 1291.3 65845 525 57785
39325 12955 2805 16.8 16 29180 355 45 2775 64048 0.5 2.9 38513 3660 82 5

2437000 FRANCHEVILLE 130460 137458 121.88 1127.8 133730 1675 114030 71550
21560 9695 13.2 13.5 67205 1315 95 3875 76539 0.5 2.8 43170 6490 83 5

2438000 BECANCOUR 19269 19175 16.89 1135.6 18775 280 16875 11885 3985 780
11.6 11.3 8570 960 20 305 62569 0.4 3 41147 640 84 5

2439000 ARTHABASKA 58224 60257 31.79 1895.7 59235 790 52000 34380 11435
2570 11.1 9.3 29675 2240 95 910 71531 0.5 3 40591 2355 85 5

2440000 ASBESTOS 16223 15381 19.92 772.3 14515 225 13535 10030 3300 505 12.6
13.1 6440 655 570 180 55871 0.4 3 38042 605 86 5

2441000 LE HAUT-SAINT-FRANCOIS 20801 20769 8.8 2359.1 17760 380 18335
12860 4700 580 11.6 11.8 9580 965 45 330 58582 0.4 3 36394 850 87 5

2442000 LE VAL-SAINT-FRANCOIS 32181 32304 23.13 1396.9 29180 480 28020
18935 5630 1120 11 10.6 15870 980 30 505 76639 0.5 3 41292 965 88 5

2443000 SHERBROOKE 118791 127224 301.76 421.6 114605 4945 97635 55205
17265 12395 10.7 11.5 65430 670 35 4150 92178 0.5 2.8 43257 6065 89 5

2444000 COATICOOK 15276 15758 13.56 1162.1 13975 510 13545 9375 3080 520 7.9
10 7570 1075 30 305 71268 0.5 3.2 37988 585 90 5

2445000 MEMPHREMAGOG 33701 35984 28.09 1281 27095 1625 30095 19275 6525
2505 12.3 11.2 16910 410 195 720 102392 0.3 3 40489 1190 91 5

2446000 BROME-MISSIQUOI 43862 45257 29.24 1547.9 32065 2015 37805 25085
7465 2275 9.7 10.6 21920 1835 95 1060 90969 0.3 3 42239 1455 92 5

2447000 LA HAUT-YAMASKA 64803 73351 97.18 754.8 68050 1560 59225 34200
12600 3380 10 12.5 37705 1330 30 1375 92299 0.5 3 42925 2880 93 5

2448000 ACTON 14189 14613 25.28 578 14190 290 12595 8945 3405 270 10 11.1 7090
1090 35 175 68840 0.5 3.2 38867 525 94 5

2449000 DRUMMOND 75192 79654 48.95 1627.2 77390 1195 66420 43305 14515
3595 10 11.6 38965 2105 120 1490 73156 0.5 3 40338 3285 95 5

2450000 NICOLET-YAMASKA 24243 23897 23.86 1001.4 23385 560 20205 14755
4505 1155 10.6 11 11110 1520 30 550 60457 0.5 3 38551 800 96 5

2451000 MASKINONGE 23659 23802 11.89 2002.1 23515 175 20645 15345 6200 600
13.1 14 11530 1315 15 430 60947 0.4 2.9 37605 925 97 5

2452000 D'AUTRAY 32824 35727 30.07 1188 34910 545 30615 20195 7870 900 11.6
12.6 16680 1735 65 585 73442 0.5 3 39026 1370 98 5

2453000 LE BAS-RICHELIEU 53530 53909 90.85 593.4 52680 965 46850 33020 9825
1995 15.8 14.1 25750 1005 220 970 72211 0.5 3 43831 2200 99 5

2454000 LE MASKOUTAINS 73224 76828 59.45 1292.4 75300 1370 63035 40500
14340 3415 8.5 9.5 39155 3495 75 1720 90438 0.5 3 43177 2765 100 5

2455000 ROUVILLE 28700 31370 58.1 539.9 30135 655 26500 16380 5675 1115 8.7
11.6 16580 1740 40 605 91638 0.5 3 45163 890 101 5

2456000 LE HAUT-RICHELIEU 82401 92889 99.85 930.3 85530 3200 75785 45115
14895 5250 9.6 11.8 48145 1885 230 5160 92504 0.5 3 43826 3360 102 5

2457000 LA VALLEE-DU-RICHELIEU 94871 105032 189.11 555.4 92595 4985 90070
55525 9460 9970 7.5 10.3 57275 1070 135 3800 122091 0.3 3 57805 2795 103 5

2458000 CHAMPLAIN 293054 312734 1920.97 162.8 241590 37590 255710 144175
38780 30155 10.5 10.6 171845 545 125 11470 127923 0.5 3 51403 14645 104 5

2459000 LAJAMMERAIS 72179 85720 206.9 414.3 81870 2220 73945 44420 7650
8290 6.8 9.5 48575 485 130 3435 123515 0.5 3.1 60771 1715 105 5

2460000 L'ASSOMPTION 73717 91537 361.66 253.1 88215 1825 77155 42900 10295
5470 7.6 11 49285 605 170 3515 111791 0.5 3 52287 2860 106 5

2461000 JOLIETTE 44488 48303 150.52 320.9 47250 530 39900 26155 7540 2705 10.5
8.8 23960 990 45 1050 88529 0.5 3 44560 1855 107 5

2462000 MATAWINIE 30607 35253 3.28 10740.3 30820 860 29815 19115 8065 1120
17.5 16.2 15365 660 40 750 71700 0.4 2.9 36278 1505 108 5

2463000 MONTCALM 28610 32872 45.67 719.7 31785 560 27590 17985 7440 790 12
12.5 15190 1385 55 580 74981 0.5 3 37475 1280 109 5

2464000 LES MOULINS 68768 91156 345.81 263.6 86020 2320 77020 41090 10980
3435 8.7 12 48795 490 80 2885 103933 0.5 3.1 48265 3610 110 5

2465000 LAVAL 284164 314398 1281.69 245.3 239910 41740 265630 159505 40270
24685 9.5 10.5 172140 1055 185 10015 123608 0.5 3 51582 12510 111 5

2466000 COMMUNAUTE URBAINE DE MONTREAL 1752582 1775871 3598.52
493.5 975035 411865 1416650 823935 286205 229565 13.5 12.6 936995 2360 870
45880 175454 0.5 2.8 48855 108655 112 5

2467000 ROUSSILION 99403 118355 280.79 421.5 94900 6845 101190 60275 13685
6870 8.8 10.6 63915 765 75 3345 106303 0.5 3 51476 3510 113 5

2468000 LES JARDINS-DE-NAPIERVILLE 20391 21977 27.57 797.2 19275 795
18640 13080 4565 695 8 12.7 11475 2180 65 415 88630 0.5 3 42540 685 114 5

2469000 LE HAUT-SAINT-LAURENT 21307 21864 18.69 1169.9 13700 1060 19190
13420 4250 760 10.8 14.1 10480 1815 125 280 82754 0.4 3 40718 930 115 5

2470000 BEAUHAMOIS-SALABERRY 57742 59785 129.49 461.7 56795 845 49960
32265 11635 2195 14.3 12.7 29380 1115 245 1100 85277 0.5 2.9 43415 2650 116 5

2471000 VAUDREUIL-SOULANGES 69766 84503 99.16 852.2 62675 5105 71445
40905 9945 6075 8.3 9 45610 1450 95 2040 126761 0.3 3 53014 2120 117 5

2472000 DEUX-MONTAGNES 59321 71218 295.14 241.3 62450 2150 59405 34180
9700 3090 10.2 12.1 37105 815 140 2065 101376 0.5 3 47168 3255 118 5

2473000 THERESE DE BLAINVILLE 79744 104693 512.2 204.4 93315 4775 85470
43335 10830 7605 8.6 11 56615 590 25 3325 126637 0.5 3 53026 3775 119 5

2474000 MIRABEL 13875 17971 36.51 492.2 17220 275 14745 8620 2810 645 9.2 12.7
9650 885 50 400 97068 0.5 3 45302 600 120 5

2475000 LA RIVIERE-DU-NORD 62042 73896 162.27 455.4 70405 1355 58875 33010
12690 2790 13.7 14 38055 305 140 2085 93485 0.5 2.9 42353 3265 121 5

2476000 ARGENTEUL 26738 27232 21.7 1255 20385 745 22965 15940 5535 1020 12.1
11.6 12945 595 80 615 76694 0.4 2.9 39582 1130 122 5

2477000 LES PAYS-D'EN-HAUT 18857 23088 31.24 739 18980 1255 19220 11350
3265 2260 11.3 13.6 12075 100 50 660 130844 0.3 2.7 49477 675 123 5

2478000 LES LAURENTIDES 28594 31580 12.87 2454.7 28445 945 26215 16095 6385
1560 14.6 16 15545 340 40 895 89914 0.3 2.7 37636 1280 124 5

2479000 ANTOINE-LABELLE 30906 32019 2.15 14905.7 31095 290 26565 17575
7465 995 19.6 15.6 14460 490 240 1020 63042 0.5 3 36380 1420 125 5

2480000 PAPINEAU 18790 19526 6.63 2946.5 17840 260 16560 12050 4770 585 14.5
14.5 9180 620 0 750 74015 0.4 2.9 36539 800 126 5

2481000 COMMUNAUTE URBAINE DE L'OUTAOUAIS 182604 201536 428.44
470.4 165300 11190 155725 86895 23100 19330 8.3 8 113420 645 130 27015 104548
0.5 3 52699 8270 127 5

2482000 LES COLLINES-DE-L'OUTAOUIS 20355 28894 14.91 1938.3 19060 1050
24475 15420 3910 2675 10.1 9 15735 460 20 3305 110770 0.3 3 52509 685 128 5

2483000 LA VALLEE-DE-LA-GATINEAU 19485 18706 1.48 12621.6 15665 235
16160 11630 5135 595 20.7 20.1 8435 210 15 815 59386 0.4 2.9 36244 1020 129 5

2484000 PONTIAC 15100 15111 0.97 15597 5645 310 13280 10140 3490 470 22.2 15.8
6810 510 80 400 64153 0.4 3 37831 710 130 5

2485000 TEMISCAMINGUE 17332 17381 0.92 18948.6 14690 235 14625 10045 3315
580 18.6 16.2 7855 635 165 440 57956 0.5 3.2 40466 675 131 5

2486000 ROUYN-NORANDA 39579 42033 7.01 5998.7 39545 660 34275 20080 7005
2440 16.5 14 21545 165 2855 1575 83942 0.5 3 48443 1285 132 5

2487000 ABITIBI-OUEST 24293 24109 7.16 3367.1 23705 150 20460 14520 5420 660
21 18.2 10770 450 785 380 56652 0.5 3.2 40858 950 133 5

2488000 ABITIBI 25222 25334 3.17 8003.7 24530 175 20970 13945 4835 1015 19.8
14.6 12125 505 510 635 68215 0.5 3.2 43426 710 134 5

2489000 VALLEE-DE-L'OR 40344 43121 1.59 27165.4 39360 635 34210 19850 8140
1915 18.1 17.6 21190 145 2925 970 80322 0.5 3 44653 1600 135 5

2490000 LE HAUT-SAINT-AURICE 16389 16272 0.58 28026.3 13800 85 14120
9625 3275 390 13.8 13 7270 65 10 510 65242 0.5 3 41815 490 136 5

2491000 LE DOMAIN-DU-ROY 33307 33239 1.78 18652.5 32555 160 28715 20370
5485 1465 22.2 15.6 15400 600 35 920 61826 0.5 3.2 41599 1070 137 5

2492000 MARIA-CHAPDELAIN 28920 28164 0.73 38673.8 27870 100 24595 17720
5215 820 19.2 19.8 12370 765 55 670 56028 0.5 3.4 38351 1170 138 5

2493000 LAC-SAINT-JEAN-EST 52413 51963 19.01 2733 51405 210 45330 31640
7740 2265 15.8 16 23400 1175 190 1200 62514 0.5 3.2 42063 1875 139 5

2494000 LE FJORD-DU-SAGUENAY 170817 172793 3.8 45502.6 169545 1175
148035 98815 23005 10450 13.2 15 79925 1190 285 6520 73754 0.5 3 44983 7180 140 5

2495000 LA HAUTE-COTE-NORD 14263 13541 1.02 13315.8 13400 60 11990 8890
3015 450 29.8 21.7 6085 120 80 445 43328 0.5 3.2 36235 785 141 5

2496000 MANICOUAGAN 36369 36108 1.14 31780.3 33455 255 30050 20735 5345
1590 10.1 14.6 18535 80 95 1560 69065 0.5 3 49666 1070 142 5

2497000 SEPT-RIVIERES-CANIAPISCAU 40891 40730 0.45 90514 36090 600 33775
21480 6335 1785 12.1 15.8 21090 25 3420 1850 61386 0.5 3 48578 1410 143 5

2498000 MINGANIE-COTE-NORD-DU-GOLFE-SAINT-LAURENT 13075 12845
0.13 100131.7 6880 20 11445 9315 3670 310 38 34.9 6160 10 310 550 43622 0.5 3.5
40338 395 144 5

2499000 TERRITOIRE NORDIQUE 36112 36310 0.05 741906.5 20160 285 29645
17295 7040 1080 14.6 15.8 15940 50 1905 2130 49445 0.6 3.6 47185 760 145 5

3501000 STORMONT, DUNDAS AND GLENGARRY UNITED COUNTIES 102262
107841 32.66 3301.8 25105 7310 89250 55705 12110 5250 8.1 8.5 54400 4090 120 4040
105567 0.3 3 46184 3560 146 6

3502000 PRESCOTT AND RUSSELL UNITED COUNTIES 57620 67183 33.54 2003.3
45980 2465 55060 32845 8705 3475 6.5 6.5 34860 2370 55 4455 112945 0.3 3 50951
1555 147 6

3506000 OTTAWA-CARLETON REGIONAL MUNICIPALITY 606639 678147
245.98 2756.9 109860 122085 528500 277310 35230 124200 6.8 7 395005 3725 260
94440 181468 0.3 3 64815 19800 148 6

3507000 LEEDS AND GRENVILLE UNITED COUNTIES 84582 90235 26.62 3389.9
1935 8000 75475 47035 6800 5870 6.5 6.5 46880 2355 100 4220 124958 0.3 2.8 48787
2085 149 6

3509000 LANARK COUNTY 49649 54803 17.89 3064 1175 3765 44770 27360 4425
3755 6.8 7.8 27615 1220 95 2775 123408 0.3 3 50930 1100 150 6

3510000 FRONTENAC COUNTY 115221 129089 33.8 3819.7 3335 17005 99175
55515 8265 16455 7.2 8 69735 1070 150 9850 160626 0.3 2.8 53277 3530 151 6

3511000 LENNOX AND ADDINGTON COUNTY 34354 37243 13.11 2840.7 430 2570
30605 18720 3060 2085 10.2 8.1 18705 970 35 1810 137508 0.3 3 47729 900 152 6

3512000 HASTINGS COUNTY 109352 116434 19.51 5967.3 2670 9135 94275 56545
10625 6030 9.5 10.5 58535 1825 140 6935 137634 0.3 3 46483 3510 153 6

3513000 PRINCE EDWARD COUNTY 22427 23763 22.67 1048.3 180 2230 19670
13080 2370 1270 6.7 7.2 11815 1165 10 965 162578 0.3 2.9 49255 545 154 6

3514000 NORTHUMBERLAND COUNTY 67704 78224 37.12 2107.6 850 8325 64575
37205 6205 4155 7.5 8.2 39760 2415 190 2765 164538 0.3 3 50321 1475 155 6

3515000 PETERBOROUGH COUNTY 105056 119992 30.33 3956.1 950 10980 98615
57505 9875 8250 8.5 8.7 59380 1940 365 3135 164946 0.3 3 48346 3495 156 6

3516000 VICTORIA COUNTY 52599 63332 20.65 3066.8 520 5705 53825 31210 5355
2900 8.1 9.6 30815 2100 185 1540 171187 0.3 3 47983 1350 157 6

3518000 DURHAM REGIONAL MUNICIPALITY 326179 409070 164.31 2489.6 7045
76960 332290 164660 23205 29065 7 8.1 227795 4415 510 14715 208740 0.3 3.1 60937
8370 158 6

3519000 YORK REGIONAL MUNICIPALITY 350602 504981 287.64 1755.6 5180
161015 423670 202020 35380 66685 5.8 7 288655 3820 555 14415 323351 0.3 3.2
74142 8340 159 6

3520000 TORONTO REGIONAL MUNICIPALITY 2192721 2275771 3612.33 630
28805 957370 1844535 1143080 248180 336100 10 9 1287950 3115 2025 72325 287045
0.5 3 60319 95980 160 6

3521000 PEEL REGIONAL MUNICIPALITY 592154 732798 598.2 1225 10440
264205 595275 296165 49185 74510 7.3 8.1 433460 3325 785 21700 247937 0.5 3.1
63521 16760 161 6

3522000 DUFFERIN COUNTY 32650 39897 26.77 1490.3 400 5240 33525 17115 2600
2500 6.7 6.7 22430 1495 175 1165 207106 0.3 3.2 57482 700 162 6

3523000 WELLINGTON COUNTY 139447 159609 60.02 2659.3 1645 26125 130375
73505 12945 16465 6.5 7.1 88810 5855 200 4800 195602 0.3 3 56886 2905 163 6

3524000 HALTON REGIONAL MUNICIPALITY 271389 313136 326.66 958.6 5345
70985 263870 147190 13860 40305 5.7 6.2 183155 2365 790 10255 253425 0.3 3 73287
4690 164 6

3525000 HAMILTON-WENTWORTH REGIONAL MUNICIPALITY 423398 451665
405.85 1112.9 5885 108715 377635 231030 48690 35190 10.6 8.7 236880 4040 530
11455 180861 0.3 3 52267 18295 165 6

3526000 NIAGARA REGIONAL MUNICIPALITY 370132 393936 212.85 1850.8
13560 73205 331805 209150 39460 25725 9.1 9.2 202260 7430 505 11160 154822 0.3 3
50590 11770 166 6

3528000 HALDIMAND-NORFOLK REGIONAL MUNICIPALITY 90121 98707 33.91
2910.8 965 12085 84520 53260 11525 4060 6.6 7 51635 8010 330 2540 139448 0.3 3
49329 2010 167 6

3529000 BRANT COUNTY 106267 110806 101.55 1091.2 1065 15060 92100 53635
10355 6185 10 8.3 57470 2490 105 2630 152672 0.3 3 49588 3175 168 6

3530000 WATERLOO REGIONAL MUNICIPALITY 329404 377762 277.83 1359.7
4745 77750 307555 169230 35330 33965 8.6 8.6 212455 4735 255 8860 176388 0.3 3
55103 9845 169 6

3531000 PERTH COUNTY 66597 69976 31.95 2190.1 315 6425 59170 38405 8265
3750 5.5 5 37935 5190 30 1395 149117 0.3 3.2 51158 1350 170 6

3532000 OXFORD COUNTY 85364 92888 45.71 2032.2 845 11135 76975 46655
10375 4120 6.5 7.5 49195 6035 115 1980 143566 0.3 3 52133 1945 171 6

3534000 ELGIN COUNTY 70335 75423 40.12 1879.9 615 10315 62475 38575 8095
3065 9.1 9.2 39360 4575 85 1870 128453 0.3 3 48583 1715 172 6

3536000 KENT COUNTY 106732 109943 44.08 2494.3 3625 11865 91705 58940
12610 5390 9.3 8.8 57440 5800 305 3040 105993 0.3 3 48074 3055 173 6

3537000 ESSEX COUNTY 316362 327365 175.86 1861.5 14590 66040 274850 175335
34735 24245 11.3 10.7 166685 5145 425 7440 128200 0.3 3 53042 10295 174 6

3538000 LAMBTON COUNTY 124592 128943 43.03 2996.5 2760 16700 108570
66680 9690 8370 9.1 8 66990 4065 860 2940 132634 0.3 3 55263 3050 175 6

3539000 MIDDLESEX COUNTY 332471 372274 110.75 3361.3 4090 70575 296045
160210 26275 40170 9 7.1 207310 6875 275 10380 163255 0.3 3 55233 10685 176 6

3540000 HURON COUNTY 55996 59065 17.36 3402.7 385 4995 50635 33540 6505
2600 5.7 6.3 30255 4970 425 1620 116838 0.3 3.2 48209 1010 177 6

3541000 BRUCE COUNTY 58848 65268 16.12 4048.4 615 5130 55560 35330 6725
3780 6 7.3 33070 3705 150 1625 137318 0.3 3 49447 1380 178 6

3542000 GREY COUNTY 74759 84071 18.66 4505 435 6800 70415 41920 9045 4695
8.3 8.5 43235 4215 180 2290 154015 0.3 3 45660 1980 179 6

3543000 SIMCOE COUNTY 238408 288684 59.62 4842.3 8050 34315 230720 123255
21820 16740 8 8.7 153780 5295 420 13330 190312 0.3 3 53049 5835 180 6

3544000 MUSKOKA DISTRICT MUNICIPALITY 40235 48005 11.9 4035.2 565 4000
39535 23055 4095 2965 8.5 8.1 23995 290 65 1655 174002 0.3 2.9 45877 1135 181 6

3546000 HALIBURTON COUNTY 11961 14421 3.46 4168.7 110 1190 12155 7715
1850 635 14 10 6590 70 45 440 153766 0.3 2.7 39273 365 182 6

3547000 RENFREW COUNTY 88965 91685 11.99 7645.6 4215 5170 75185 48850
11085 5015 7.7 9.1 45900 1830 155 7040 101599 0.3 3 45578 2325 183 6

3548000 NIPISSING DISTRICT 79004 84723 4.7 18011.5 21310 4210 67755 39655
8950 4965 9.6 9.7 41215 445 385 5225 122756 0.3 3 45905 3155 184 6

3549000 PARRY SOUND DISTRICT 33828 38423 3.82 10056.5 770 2955 32280
20400 4535 1790 9 9 17945 460 100 1560 139762 0.3 2.9 41602 1310 185 6

3551000 MANITOULIN DISTRICT 9823 11192 3.04 3678.9 140 505 9615 6825 1570
540 9.5 8.3 4950 360 125 665 90600 0.3 3 37980 225 186 6

3552000 SUDBURY DISTRICT 25771 26178 0.6 43275 7585 905 22200 14780 3820
960 11.3 10.2 12735 330 495 1115 88216 0.5 3 44940 980 187 6

3553000 SUDBURY REGIONAL MUNICIPALITY 152476 161210 61.84 2607 43810
12960 132550 82405 18520 11240 8.1 9 82810 345 8205 8245 122107 0.5 3 54373 5265
188 6

3554000 TIMISKAMING DISTRICT 40307 38983 3.07 12705.3 9240 2020 32470
21810 5780 1745 11.1 10.1 18315 915 1305 1520 74565 0.4 3 44046 1490 189 6

3556000 COCHRANE DISTRICT 93712 93917 0.64 145618 42310 4260 77300 48430
13175 4035 12.7 11 45510 460 4775 3525 95794 0.5 3.1 49611 2925 190 6

3557000 ALGOMA DISTRICT 131841 127269 2.49 51206.8 10270 13720 106390
68720 13680 7630 11.5 11.1 61550 645 3055 4825 100799 0.3 3 46757 4720 191 6

3558000 THUNDER BAY DISTRICT 155673 158810 1.45 109564.2 6840 18600
132310 83945 16185 10845 10.3 8.2 83920 830 2440 8880 110879 0.3 3 55067 3915 192
6

3559000 RAINY RIVER DISTRICT 22871 22997 1.37 16817.1 395 2000 19355 12900
2340 1035 10.3 9.3 11285 425 105 1300 83162 0.5 3.2 48392 475 193 6

3560000 KENORA DISTRICT 52834 58748 0.15 396871 1470 4105 47710 29375 8460
2810 8.8 7.6 28790 240 1345 4580 97876 0.5 3.1 47650 925 194 6

4601000 DIVISION NO. 1 16262 16023 1.09 14725.2 1185 1385 13815 9925 2550 885
8.8 6.5 8015 1040 315 510 66770 0.4 3 44457 390 195 7

4602000 DIVISION NO. 2 40368 44222 10.37 4265.6 8045 3385 37725 24375 6840
1750 5.5 6.1 22555 3575 75 990 81552 0.5 3.5 44457 1000 196 7

4603000 DIVISION NO. 3 38422 38770 7.56 5127.7 1435 4075 32480 22545 8170 1580
4 4.5 18555 4305 50 850 68654 0.4 3.4 40262 1015 197 7

4604000 DIVISION NO. 4 11469 10549 2.39 4419.3 1130 390 9050 7230 1580 305 2
3.9 5380 2505 0 190 41844 0.4 3.2 34986 410 198 7

4605000 DIVISION NO. 5 16495 15026 1.87 8022.2 265 840 13105 9960 2155 525 1.7
3.2 7565 2910 70 410 44316 0.4 3 35286 630 199 7

4606000 DIVISION NO. 6 11176 10357 2.71 3814.9 170 420 8900 6575 1430 345 5 4.8
5045 1670 105 275 57199 0.4 3 33592 310 200 7

4607000 DIVISION NO. 7 57112 56389 10.1 5582.2 885 3540 43270 27465 4880 3705
5.9 8 30025 3125 60 3460 72419 0.4 3 42887 2165 201 7

4608000 DIVISION NO. 8 13226 14586 2.58 5653.3 975 670 11915 9380 3250 575 3.5
3.7 6850 3010 20 355 53321 0.5 3.2 34582 495 202 7

4609000 DIVISION NO. 9 23236 23711 8.24 2877.8 1625 1225 17785 11760 2935 1070
6 6.6 11945 2130 0 2015 69336 0.4 3 42643 720 203 7

4610000 DIVISION NO. 10 7334 8012 4.37 1832.8 880 450 6115 4300 845 440 3 3
4825 1655 15 275 100729 0.5 3.2 53915 95 204 7

4611000 DIVISION NO. 11 594551 616790 1079.25 571.5 25775 110250 488885
289845 52565 64425 9.3 8.1 332225 1860 505 28985 95114 0.3 3 49301 28825 205 7

4612000 DIVISION NO. 12 15859 17383 9.71 1789.3 360 1255 15190 10695 2085 1100
5.5 5.5 9575 1085 115 885 104702 0.5 3.2 47835 335 206 7

4613000 DIVISION NO. 13 33619 37204 22.63 1643.7 480 3340 31945 21640 3385
2210 8.2 7.2 20675 950 150 1915 117404 0.3 3 52879 560 207 7

4614000 DIVISION NO. 14 14713 15701 5.73 2740.9 250 745 12905 8610 1525 645 4.5
5.5 8440 1725 40 590 89960 0.5 3.2 46812 335 208 7

4615000 DIVISION NO. 15 23818 22556 2.47 9115.4 425 860 19530 14865 3775 960 5
4 11275 3770 75 710 49184 0.4 3 37661 715 209 7

4616000 DIVISION NO. 16 10577 10028 2.13 4711.8 175 420 8760 6690 2135 385 5.4
5.6 4880 1690 70 275 47826 0.4 3 34902 495 210 7

4617000 DIVISION NO. 17 26522 24897 1.87 13340.2 1060 960 21425 16340 5040 950
6 5.6 11920 3125 110 1000 51257 0.4 3 36199 965 211 7

4618000 DIVISION NO. 18 22005 21637 1.95 11103.8 700 1110 18620 13490 4125 700
7.5 9 10615 2530 85 905 57568 0.5 3 35091 865 212 7

4619000 DIVISION NO. 19 9125 12428 0.2 61609.2 50 60 10365 6665 3155 195 27.3
19.6 3425 200 40 1050 36798 0.6 3.7 21392 390 213 7

4620000 DIVISION NO. 20 12256 11510 1.16 9895 130 825 10055 7430 2505 425 6.6
6.6 5520 1555 90 295 46526 0.4 3 34323 625 214 7

4621000 DIVISION NO. 21 24068 23283 0.51 46025.1 395 960 18565 10920 2665 1075
12.6 12.1 11560 125 1645 1325 62608 0.5 3.2 49123 680 215 7

4622000 DIVISION NO. 22 30544 32101 0.34 94966 355 1225 24870 13485 5135 1080
16.3 16.5 13050 15 1980 2245 76741 0.6 3.7 42689 730 216 7

4623000 DIVISION NO. 23 10259 8779 0.04 233870 130 210 6570 3325 1445 305 15.8
17.3 3730 10 510 540 25566 0.6 3.6 47242 180 217 7

4701000 DIVISION NO. 1 33813 32181 2.15 14967.8 985 1295 27855 20515 3565 1085
3 5.4 16740 4550 1385 735 57760 0.4 3.2 45747 945 218 8

4702000 DIVISION NO. 2 26491 24186 1.43 16923.9 325 1125 20960 15820 2875 850
3 5.6 12525 4080 575 500 59160 0.4 3 43128 670 219 8

4703000 DIVISION NO. 3 19392 17628 0.95 18462.4 1940 705 15700 12680 2095 615
1.7 4.9 9595 4475 260 275 45471 0.4 3.2 40584 660 220 8

4704000 DIVISION NO. 4 14058 12793 0.6 21202.9 320 560 10610 7945 1895 335 3 5
7060 3320 150 355 49586 0.4 3 40853 455 221 8

4705000 DIVISION NO. 5 40315 37399 2.57 14554.5 370 1395 33060 25905 6420 1095
4 5.8 18155 5785 1365 850 49733 0.4 3 39701 1300 222 8

4706000 DIVISION NO. 6 213800 218612 12.56 17411.9 2885 16920 174935 108325
17480 19950 6.8 7.2 117950 7485 860 12410 79527 0.3 3 50868 7505 223 8

4707000 DIVISION NO. 7 53706 50984 2.68 18995.6 940 2840 41145 27405 5410 2255
5.1 6.1 25450 5110 220 2670 61366 0.4 3 42143 2065 224 8

4708000 DIVISION NO. 8 35723 32569 1.46 22239 290 1605 27465 19875 3615 1395
3.4 5.4 17680 5665 555 805 60392 0.4 3 43028 1110 225 8

4709000 DIVISION NO. 9 43455 40755 2.7 15115.8 150 1785 35255 26635 8590 1405
6.3 6.4 20025 5660 115 1020 57329 0.4 3 37047 1580 226 8

4710000 DIVISION NO. 10 24487 22325 1.84 12116.6 90 845 19740 16165 4765 610
4.5 5.6 10860 4380 125 560 44650 0.4 3.2 35249 975 227 8

4711000 DIVISION NO. 11 217231 224832 13.35 16844.2 3515 17870 174135 102000
16900 23225 8 8.6 120355 7520 2880 7505 82884 0.3 3 47966 8990 228 8

4712000 DIVISION NO. 12 25867 24521 1.72 14277.8 380 1015 20740 15070 2770
1060 4.5 6.9 12725 4320 280 595 59336 0.4 3.2 43819 740 229 8

4713000 DIVISION NO. 13 28656 25388 1.48 17110.8 190 1025 22095 16290 2950
1040 2.2 4.6 13115 5370 830 460 56149 0.4 3.2 42657 820 230 8

4714000 DIVISION NO. 14 46932 42618 1.27 33647.9 950 1855 37165 28780 7965
1435 6.8 7.4 20885 6705 150 1175 47906 0.4 3 36225 1610 231 8

4715000 DIVISION NO. 15 82258 80196 4.14 19348.2 4015 3350 64510 45445 11730
3715 8.3 8.6 38695 7440 440 3080 63244 0.4 3.2 41145 3445 232 8

4716000 DIVISION NO. 16 40314 38372 1.72 22303.9 1365 1800 31810 22670 6440
1630 7.5 8.3 18210 4890 140 1195 57279 0.4 3.2 36423 1790 233 8

4717000 DIVISION NO. 17 37775 36834 1.66 22159.5 720 1515 30380 20540 5490
1285 8.1 8.8 17720 4340 905 1060 64803 0.5 3.4 38047 1225 234 8

4718000 DIVISION NO. 18 25340 26735 0.11 252430 125 300 20875 11955 5840 715
25.2 18.2 8545 45 755 2005 51543 0.8 4 31021 1290 235 8

4801000 DIVISION NO. 1 56592 58179 2.83 20532.1 490 5250 44835 28075 5950 2725
7 8.5 30340 3890 1430 2420 85003 0.3 3 45875 2095 236 9

4802000 DIVISION NO. 2 115739 119110 6.81 17489.9 775 15445 90875 55155 9210
7955 6 8.2 62355 8975 1990 4320 87947 0.3 3.1 46563 3760 237 9

4803000 DIVISION NO. 3 34970 36149 2.94 12279.6 225 2600 28110 19435 3030 1840
8.3 8.3 17035 4250 410 1460 71831 0.4 3.4 41819 855 238 9

4804000 DIVISION NO. 4 12376 11947 0.55 21606.8 55 805 9860 7250 1085 500 2 5
6610 2785 285 280 61499 0.4 3.2 47548 260 239 9

4805000 DIVISION NO. 5 38805 39658 2.37 16714 305 2745 30225 20430 3105 1715
5.5 6.5 20740 6365 730 1615 77629 0.4 3.2 43702 1260 240 9

4806000 DIVISION NO. 6 715605 804802 64.77 12425.5 10855 156000 598430 304640
37610 98470 7.5 8.2 469265 10010 33985 24800 144269 0.3 3 58161 29495 241 9

4807000 DIVISION NO. 7 40681 39694 2.06 19223.6 400 2065 32835 23140 3350 1470
3.2 5 21445 6275 1850 1465 66612 0.4 3.2 46094 1000 242 9

4808000 DIVISION NO. 8 116611 124098 12.51 9918.9 1260 9845 92835 50835 7430
6615 6.5 7.6 67340 7825 4360 4110 91900 0.3 3 47380 4325 243 9

4809000 DIVISION NO. 9 15999 16734 0.89 18875.2 160 1030 13320 8750 1700 560
8.1 8.8 9000 1530 1045 360 79591 0.5 3.2 40843 615 244 9

4810000 DIVISION NO. 10 79745 78483 3.83 20517.9 815 4895 63475 43375 8775
3860 4.6 6 41540 9540 1655 2245 74135 0.4 3 44597 2400 245 9

4811000 DIVISION NO. 11 807521 876292 55.15 15889.9 19805 155260 661570
363580 54515 86090 8.1 8.1 491920 12055 12670 43235 116790 0.3 3 52597 36520 246
9

4812000 DIVISION NO. 12 43907 43277 4.92 8801.9 4800 2360 33395 20540 4375
2165 6.8 10.1 22645 2760 1490 5220 85745 0.5 3.2 45091 1330 247 9

4813000 DIVISION NO. 13 54973 57521 2.54 22607 1330 4805 46760 30745 6385
2265 6 7.1 30455 6840 1820 1530 74741 0.4 3 41983 1950 248 9

4814000 DIVISION NO. 14 25292 25784 1.02 25346.2 730 2025 20310 11380 1890
1000 7.8 9.2 14075 855 2010 700 85280 0.5 3.2 50147 735 249 9

4815000 DIVISION NO. 15 22794 26539 0.85 31206.7 775 3370 18820 10990 1685
2215 5.6 7 16145 240 810 1510 126558 0.3 3 50651 420 250 9

4816000 DIVISION NO. 16 48779 49490 0.4 124055.9 1960 4515 37275 18905 3395
2735 8.3 11.6 27025 880 7745 1875 81364 0.5 3.3 61874 1595 251 9

4817000 DIVISION NO. 17 48852 48962 0.26 186391.2 1050 2410 39755 24475 8090
1555 11.8 10.1 22385 3180 1630 2110 61963 0.6 3.5 40891 1825 252 9

4818000 DIVISION NO. 18 13587 13979 0.41 34401.7 310 925 10910 6635 1225 465
7.5 9.1 7475 1020 1370 600 66362 0.5 3.2 50386 415 253 9

4819000 DIVISION NO. 19 72997 74855 3.75 19947.5 4720 5160 56365 32790 5705
3725 6.6 8.1 41465 5025 2390 2615 78046 0.4 3.2 47806 2305 254 9

5901000 EAST KOOTENAY REGIONAL DISTRICT 53089 52368 1.85 28344.3 795
5900 42340 26620 3635 2235 9.1 12.1 27300 550 3885 1455 74245 0.4 3 49391 1210
255 10

5903000 CENTRAL KOOTENAY REGIONAL DISTRICT 49110 51073 2.2 23236.6
555 6435 40920 26900 4805 2785 14.5 13.1 24290 735 445 1460 82851 0.4 2.9 41485
1720 256 10

5905000 KOOTENAY BOUNDARY REGIONAL DISTRICT 30335 31194 3.95 7907
300 3935 25395 17325 2905 1620 12.3 11.3 15150 425 665 715 80465 0.4 3 46775 800
257 10

5907000 OKANAGAN-SIMIKAMEEN REGIONAL DISTRICT 59089 66701 6.41
10410.5 1150 11425 50300 27730 6435 3280 10.8 12 30555 2990 710 1825 106384 0.3
2.7 41336 1980 258 10

5909000 FRASER-CHEAM REGIONAL DISTRICT 57965 68681 6.36 10797.7 1245
10380 52065 26600 6110 2825 10.2 11.2 32970 2190 100 4065 120838 0.3 3 45152 2195
259 10

5911000 CENTRAL FRASER VALLEY REGIONAL DISTRICT 66435 87360 226.73
385.3 850 18105 64515 28670 7555 5015 8.5 11.5 43140 3790 95 2440 149512 0.3 3
49918 2480 260 10

5913000 DEWDNEY-ALOUETTE REGIONAL DISTRICT 69494 89968 28.51 3155.8
1185 13700 67725 32685 4930 3915 8.7 10.7 46160 1760 145 2765 161037 0.3 3 50315
2340 261 10

5915000 GREATER VANCOUVER REGIONAL DISTRICT 1336609 1542744 623.78
2473.2 19970 467345 1164790 596405 99310 182120 9 9.2 864830 10635 2745 43090
247831 0.3 3 57289 55280 262 10

5917000 CAPITAL REGIONAL DISTRICT 264614 299550 129.27 2317.3 4285 58105
226700 121310 13555 36195 7.7 7.6 156110 1970 235 24985 186717 0.3 2.7 52577 8040
263 10

5919000 COWICHAN VALLEY REGIONAL DISTRICT 52466 60560 17.92 3379.3
555 8875 47795 26555 4505 3560 9.5 10.3 28815 970 40 1635 122959 0.3 2.9 46680
1570 264 10

5921000 NANAIMO REGIONAL DISTRICT 82180 101736 49.84 2041.3 1355 15985
76830 38095 6210 5985 11.7 13.5 48925 675 50 2985 122946 0.3 2.7 46281 3800 265 10

5923000 ALBERNI-CLAYOQUOT REGIONAL DISTRICT 30341 31224 4.04 7738.2
605 4230 25015 15610 3045 1245 12.8 14.6 15615 210 40 945 83877 0.4 3 48647 900
266 10

5925000 COMOX-STRATHCONA REGIONAL DISTRICT 71145 82729 4.17 19851.1
1610 10660 63010 31935 4725 4500 9.8 13.6 42740 855 805 3775 107786 0.3 3 49065
2470 267 10

5927000 POWELL RIVER REGIONAL DISTRICT 18374 18477 3.62 5101 470 2825
15150 9685 1475 890 14.1 12.5 8980 130 150 290 89036 0.3 2.9 48040 570 268 10

5929000 SUNSHINE COAST REGIONAL DISTRICT 16758 20785 5.36 3879.1 275
3230 15845 8185 1015 1400 7.5 12 10030 95 85 445 151127 0.3 2.7 50600 425 269 10

5931000 SQUAMISH-LILLOOET REGIONAL DISTRICT 17891 23421 1.42 16533.8
450 3060 16930 8490 1285 1705 11.5 13.1 13570 250 65 955 150234 0.3 3 50713 510
270 10

5933000 THOMPSON-NICOLA REGIONAL DISTRICT 96805 104386 2.33 44872
1300 11775 81290 46095 7825 5290 12.5 14 54570 1900 2095 3450 84960 0.3 3 47366
3670 271 10

5935000 CENTRAL OKANAGAN REGIONAL DISTRICT 89730 111846 37.2 3006.6
1975 16710 82650 40460 9050 6405 11.8 12.7 55295 2355 510 2315 135417 0.3 2.7
46029 3515 272 10

5937000 NORTH OKANAGAN REGIONAL DISTRICT 54820 61744 7.89 7830.4 800
7855 47450 25375 5640 2920 12.5 13.3 29695 1660 100 1140 111161 0.3 2.9 43291
1925 273 10

5939000 COLUMBIA-SHUSWAP REGIONAL DISTRICT 39917 41665 1.38 30179.6
545 4500 33530 19775 3555 1815 13.5 13 20670 935 230 940 88970 0.3 2.9 42789 1240
274 10

5941000 CARIBOO REGIONAL DISTRICT 59495 61059 0.88 69168.8 840 6990
48615 28220 5435 2440 14 14.6 31685 1550 520 1725 76684 0.4 3 44809 1935 275 10

5943000 MOUNT WADDINGTON REGIONAL DISTRICT 14934 13896 0.65 21464.4
130 1445 10850 5855 920 630 10.8 14.5 7705 30 590 465 71580 0.5 3.2 52031 340 276
10

5945000 CENTRAL COAST REGIONAL DISTRICT 3120 3482 0.14 25122.2 25 240
2585 1605 400 165 21.2 15.8 1705 15 10 255 89038 0.5 3.2 36772 30 277 10

5947000 SKEENA-QUEEN CHARLOTTE REGIONAL DISTRICT 23061 23769 1.47
16139.8 360 3130 18230 9785 1830 1270 15.7 16.2 13610 30 15 1550 91349 0.5 3.2
53837 495 278 10

5949000 KITIMAT-STIKINE REGIONAL DISTRICT 39483 42053 0.44 95797 815
5935 32400 18150 3755 1930 14 15.3 21905 235 265 1870 76022 0.5 3.2 52891 960 279
10

5951000 BULKLEY-NECHAKO REGIONAL DISTRICT 37470 38343 0.53 72101.4
375 5165 30770 18220 3540 1760 13.8 11.3 19550 1070 955 1400 75748 0.5 3.4 50203
765 280 10

5953000 FRASER-FORT GEORGE REGIONAL DISTRICT 89337 90739 1.77 51196.8
1700 10175 71365 39815 6535 4355 13.6 13.3 50430 895 230 2695 81502 0.3 3.1 52572
2670 281 10

5955000 PEACE RIVER REGIONAL DISTRICT

2. Correlation Coefficient and Coefficient of Variation

Correlation Coefficient

The correlation coefficient measures the strength of the linear association between two interval/ratio scale variables. (Bivariate relationships are denoted with a small r .) Though it does not distinguish explanatory from response variables and is not affected by changes in the unit of measurement of either or both variables (Moore and McCabe, 1993).

Multiple correlation coefficients, R

$-1 \leq r \leq 1$, whereas $0 \leq R \leq 1$

(or the multiple coefficient of determination, $0 \leq R^2 \leq 1$)

the proportion of dependent variable (Y) that can be attributed to the combined effects of all the X independent variables acting together.

- for net effects (multivariate), assess R , R^2 ,
- for individual effects (bivariate) assess r , r^2 .

R - simple correlation between residuals,

R^2 - denotes the percentage of variation in the dependent variable accounted for by the independent predictor variables.

Adjusted R-Squared

An adjusted R-squared, takes the size of the sample into effect. Use when need to compare the results of models which had a differing number of observations or independent variables, or to temper the results of an analysis with suspect results due to a small number of observations.

$$\text{adjusted } R^2 = R^2 - (k - 1) / (n - k) * (1 - R^2)$$

Where:

n = # of observations,
 k = # of independent variables,

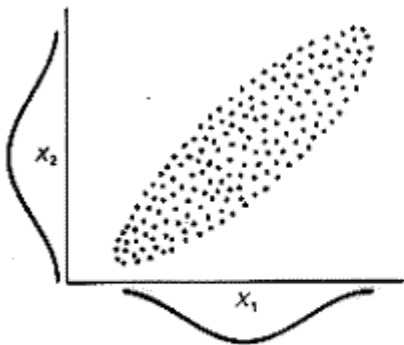
Accordingly:

smaller n , decreases R^2 value,
larger n , increases R^2 value,

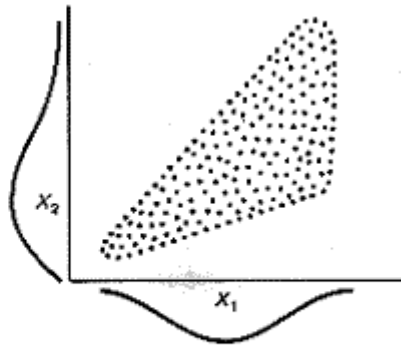
smaller k , increases R^2 value,
larger k , decreases R^2 value.

3. Homoscedasticity

Homoscedasticity is the assumption that the variability in scores for one variable is roughly the same at all values of the other variable, which is related to normality, as when normality is not met, variables are not homoscedastic.



Homoscedasticity with both variables normally distributed

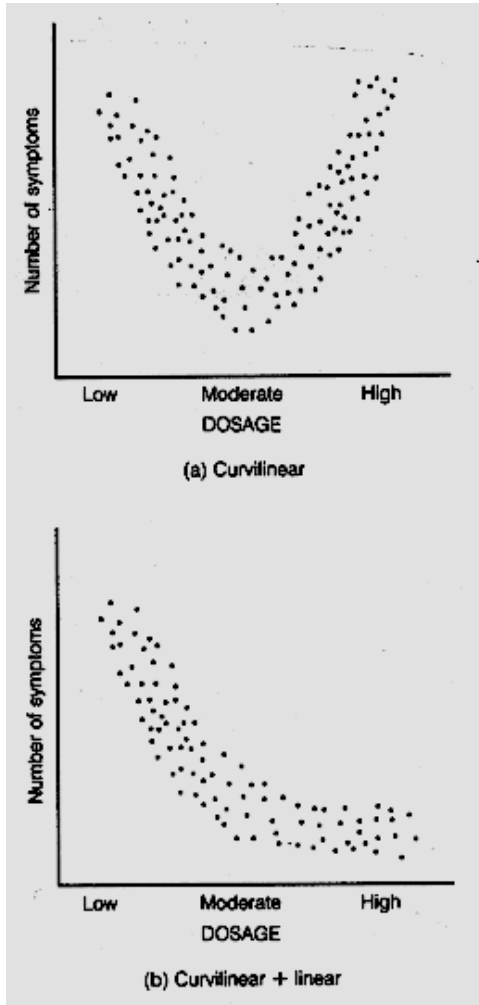


Heteroscedasticity with skewness on one variable

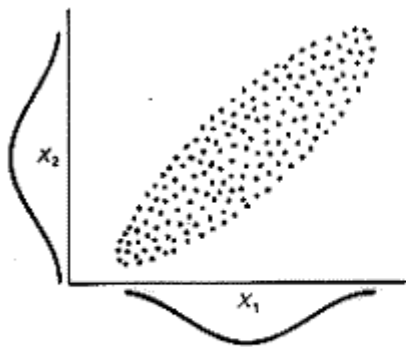
Heteroscedasticity is caused by nonnormality of one of the variables, an indirect relationship between variables, or to the effect of a data transformation. Heteroscedasticity is not fatal to an analysis, the analysis is weakened, not invalidated. Homoscedasticity is detected with scatterplots and is rectified through transformation.

4. Linearity

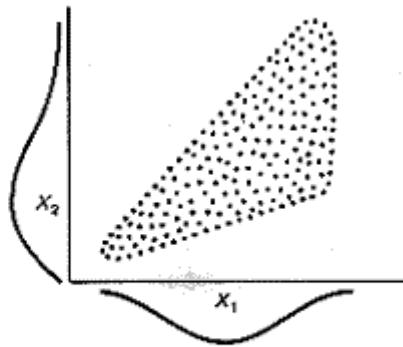
Linearity is the assumption that there is a straight line relationship between variables. Examples of non-linear distributions:



Linearity is essential for calculation of multivariate statistics due to the basis upon the general linear model, and the assumption of multivariate normality which implies that there is linearity between all pairs of variables, with significance tests based upon that assumption. Non-linearity may be diagnosed from bivariate scatterplots between pairs of variables or from a residual plot, with predicted values of the dependent variable versus the residuals. Residual plots may demonstrate: assumptions met, failure of normality, nonlinearity, and heteroscedasticity. Linearity between two variables may be assessed through observation of bivariate scatterplots. When both variables are normally distributed and linearly related, the scatterplot is oval shaped, if one of the variables is nonnormal then the scatterplot is not oval.



Homoscedasticity with both variables normally distributed



Heteroscedasticity with skewness on one variable

Refer to the homoscedasticity page for further information.

5. Multivariate General Linear Hypothesis (MGLH)

The Multivariate general linear hypothesis can estimate and test any univariate or multivariate general linear model, such as multiple regression, analysis of variance, discriminant analysis, or principle components analysis. All these procedures have their genesis in the same linear model. Linear models are based upon lines, more generally, they are based on linear planes or surfaces. Linear models are applied with wide acceptance as lines or planes often are able to describe relations among events in the real world. Accordingly, linearity of variables is very important when applying the general linear model. Linearity is the assumption of a straight line fit between variables (SYSTAT, 1992). As pairs are understood to have a linear relationship with each other, that the relationship between the variables is adequately represented by a straight line. Additivity is also important to the general linear model, as one set of variables may be predicted from another set of variables, the effect of the variables within the data set are additive in the prediction equation. The second variable in the set provides predictability to the first, the third variable provides predictability to the first two, and so on. Accordingly, in multivariate solutions, the equation which relates the set of variables is composed of a series of weighted terms added together. The assumption of linearity does not limit the use of variables with non-linear, curvilinear relationships, or multiplicative relationships. The data may be made suitable for analysis through transformation.

6. Missing Data

If the data set is large and a few random points are missing the problem is not serious, yet in a smaller data set with a non-random distribution of missing values the problem may be serious. Dealing with the problem:

- deleting cases, if a case is missing values it may be deleted. Deletion is often the default option with statistical software packages.
- estimate missing data, estimate the missing values and use these values during subsequent analysis. Estimates may be made upon, prior knowledge, inserting mean values, and using regression.
- a missing data correlation matrix may be computed. The matrix is computed using only values which are common to both variables, ie. if two cases have 15 corresponding values of 20 the correlation matrix value will be computed from those 15 common cases.
- spatially autoregressive models.
- treating missing data as data. In sociological studies it is potentially the case that a failure to respond may be indicative of some form of behavior.
- check results with and without missing data. assess the results of each analysis, if they are markedly different attempt to discern the reason for the difference. Attempt to evaluate which result more closely approximates reality (Tabachnick and Fidell, 1989).

7. Multicollinearity and Singularity

Multicollinearity and singularity are issues which are derived from the having a correlation matrix with too high of correlation between variables. Multicollinearity is when variables are highly correlated (0.90 and above), and singularity is when the variables are perfectly correlated. Multicollinearity and singularity expose the redundancy of variables and the need to remove variables from the analysis. Multicollinearity and singularity can cause both logical and statistical problems. Logically, redundant variables weaken the analysis (except in the case of factor analysis), through reduction of degrees of freedom error. Accordingly, unless doing a factor analysis, avoid variables with a bivariate correlation of greater than 0.70 in the same analysis. The statistical problems related to singularity and multicollinearity are related to matrix stability and ability for matrix inversion. With multicollinearity the effects are additive, the independent variables are inter-related, yet effecting the dependent variable differently. The high the multicollinearity, the greater the difficulty in partitioning out the individual effects of independent variables. Accordingly, the partial regression coefficients are unstable and unreliable.

Most programs appear to automatically screen for multicollinearity and singularity by computing the squared multiple correlation of a variable. The squared multiple correlation is computed where a variable is compared to all the rest of the included

variables, if the results show a high correlation the variable is multicollinear. If the variable is perfectly related to the other variables then singularity is present. Large standard errors due to multicollinearity result in both a lessened probability of rejecting the null hypothesis and wide confidence intervals.

To identify multicollinearity:

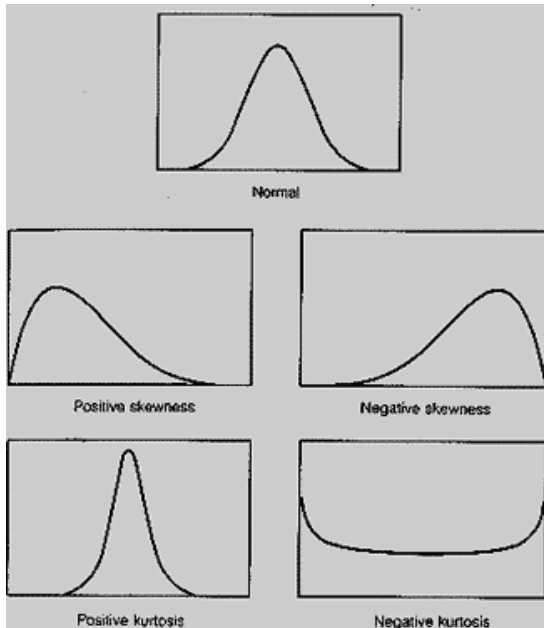
- look at pairwise relationships between variables, if r values are greater than $|0.80|$ the variables are strongly inter-related and should not be used.
- tolerance of variable, a value of near one indicates independence, if the tolerance value is close to zero, the variables are multicollinear.
- VIF, variance inflation factor, if highly collinear a high value is calculated.
- eigenvalues, view eigenvalues: no multicollinearity if eigenvalues are approximately the same size. If some are much larger than others this is indicative of the related variable loading together. Under independence the eigenvalues are normally distributed, with equal importance, yet under multicollinearity, the distribution has a few high eigenvalues and many low ones, reflecting an uneven importance of variables.
- For analysis of multicollinearity in regression analysis results some packages, such as SPSS, generate a VIF and tolerance value
- the VIF, or variance inflation factor, will reflect the presence or absence of multicollinearity. A high VIF, larger than one, the variable may be affected by multicollinearity. The VIF has a range 1 to infinity.
- tolerance has a range from zero to one. The closer the tolerance value is to zero relates a level of multicollinearity.

To Solve for Multicollinearity

- turn variables to rates,
- reduce data set, remove variables which are redundant due to a very high relationship with one another.

8. Normality

The underlying assumption of most multivariate analysis and statistical tests is the assumptions of multivariate normality. Multivariate normality is the assumption that all variables and all combinations of the variables are normally distributed. When the assumption is met the residuals are normally distributed and independent, the differences between predicted and obtained scores (the errors) are symmetrically distributed around a mean of zero and there is no pattern to the errors. Screening for normality may be undertaken in either statistical or graphical methods.



How to assess and deal with problems:

Statistically:

- examine skewness and kurtosis. when a distribution is normal both skewness and kurtosis are zero. Kurtosis is related to the peakedness of a distribution, either too peaked or too flat. skewness is related to the symmetry of the distribution, the location of the mean of the distribution, a skewed variable is a variable whose mean is not in the center of the distribution. Tests of significance for skewness and kurtosis test the obtained value against a null hypothesis of zero. Although normality of all linear combinations is desirable to ensure multivariate normality it is often not testable. Therefore, normality assessed through skewness and kurtosis of individual variables may indicate variables which may require transformation.

Graphically:

- view distributions of the data and compare to the distributions above.
- probability plots, where scores are ranked and sorted, an expected normal value is compared for the actual normal value for each case. If a distribution is normal, the points for all cases fall along the line running diagonal from lower left to upper right. Deviations from normality shift the points away from the diagonal.
- examine the residuals, plot expected values vs. obtained scores (predicted vs. actual).

If nonnormality is found in the residuals or the actual variables transformation may be considered. Univariate normality does not ensure multivariate normality, but does increase the likelihood. Transformations are recommended as a remedy for outliers, breaches in normality, non-linearity, and lack of homoscedasticity. Although

recommended, be aware of the change to the data, and the adaptation to the change which must be implemented for interpretation of results. If the scale is arbitrary interpretation will only be hindered marginally, yet if the scale is meaningful the transformation may cause confusion. Transformations are appropriate when the non-linearity is monotonic, if the distribution plot is non-monotonic rethink the variable. Sample distributions and appropriate transform to produce normality can be found on the transformations link. Remember to check the transformation for normality after application.

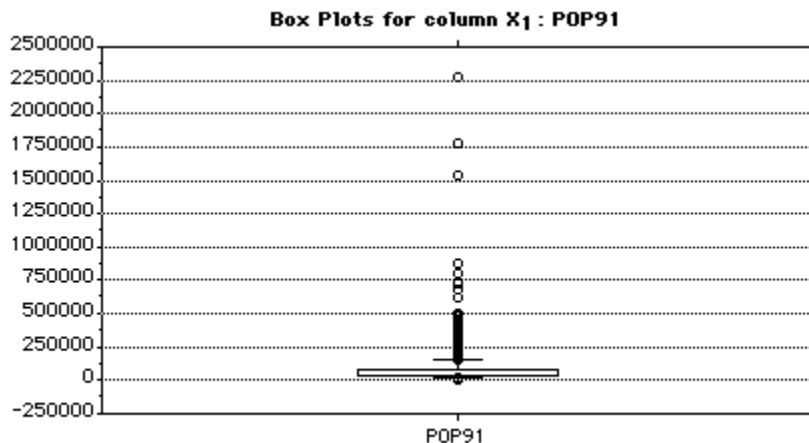
9. Orthogonality

Orthogonality relates to independence, where there should be non-association between variables, (An $R = 0$, see correlation coefficient location.).

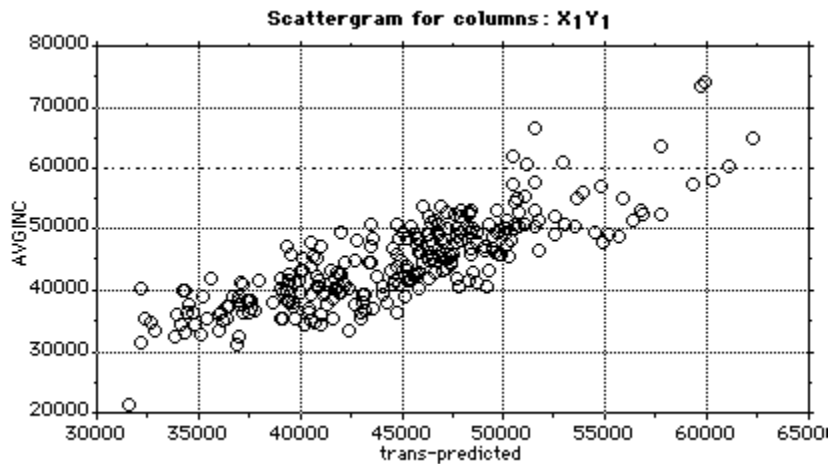
Orthogonality is perfect non-association between variables. Independence of variables is desired so that each addition of independent variable adds to the prediction of the independent variable. If the relationship between independent and dependent variables is orthogonal, the overall effect of an independent variable may be partitioned into effects on the dependent variable in an additive fashion. In orthogonal experimental designs, with random assignment, causality may be assigned to various main effects and interactions.

10. Outliers

Outliers are extreme cases on one variable, or a combination of variables, which have a strong influence on the calculation of statistics. The following box plot provides an example of outliers.



It is also possible to view for possible presence of outliers using scatterplots. Variables may be plotted both in univariate and bivariate combinations. Below is a bivariate plot of some multivariate regression results scatterplotted which reveal an outlier.



Reasons for outliers:

- incorrect data entry,
- failure to specify missing value codes,
- case not a member of intended sample population,
- actual distribution of the population has more extreme cases than a normal distribution.

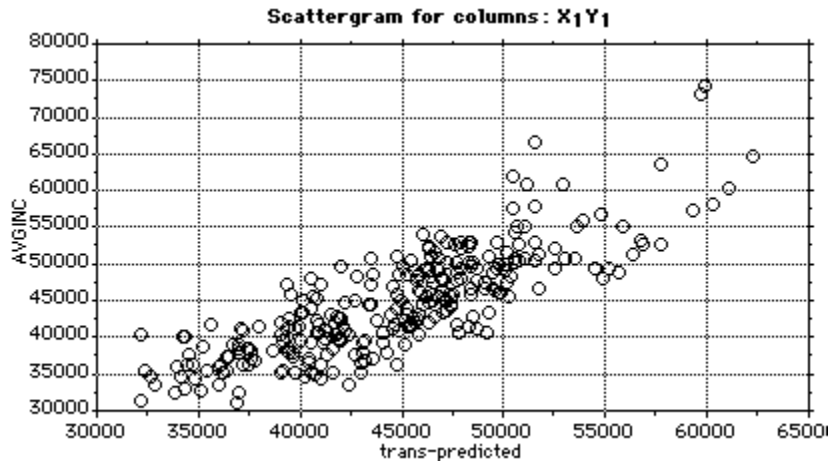
Reducing the influence of outliers:

- check the data for the case, ensure proper data entry
- check if one variable is responsible for most of the outliers, consider deletion of the variable,
- delete the case if it is not part of the population,
- if variables are from population, yet to remain in the analysis, transform the variable to reduce influence.

Graphical methods for identifying multivariate outliers:

1. Observe the residuals to identify cases for which there is a poor fit between obtained and predicted DV scores. A residuals plot will isolate cases which are outliers. Multivariate outliers will produce points outside the general swarm of points.
2. Leverage on the X axis versus residuals on the Y axis. Leverage is a distance measure, such as Mahalanobis distance. Outliers will again be exposed by lying outside the general swarm of points.

The outlier which is present above will cause problems with linearity. As a comparison the same plot is produced with the outlying value removed. (removal of outlier, case 4619000, Division 19, Province. 7 (Manitoba)). An improved linear relationship is seen with the removal of the outlier.



11. Residual Plots

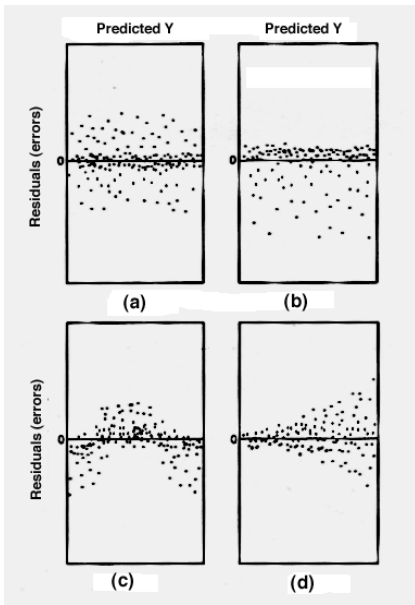
Residuals are the difference between an observed value of the response variable and the value predicted by the model (Moore and McCabe, 1993).

therefore: $\text{residual} = \text{observed } y - \text{predicted } y$

The mean of the residuals is always zero. As a result plotting of residuals enables the data to be viewed from a standard orientation point. Residual plots show the deviation from the expected value for each x value in the model.

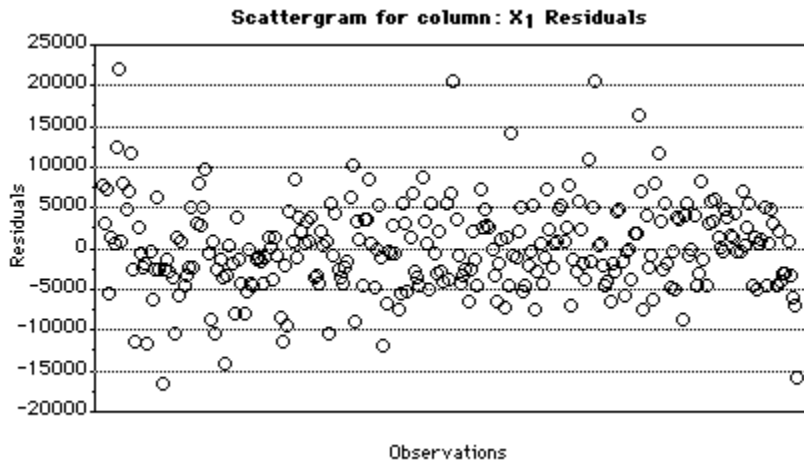
Plots of predicted values versus error (residual) values

The plots demonstrate, (a) assumptions met, (b) failure of normality, (c) nonlinearity, and (d) heteroscedasticity.

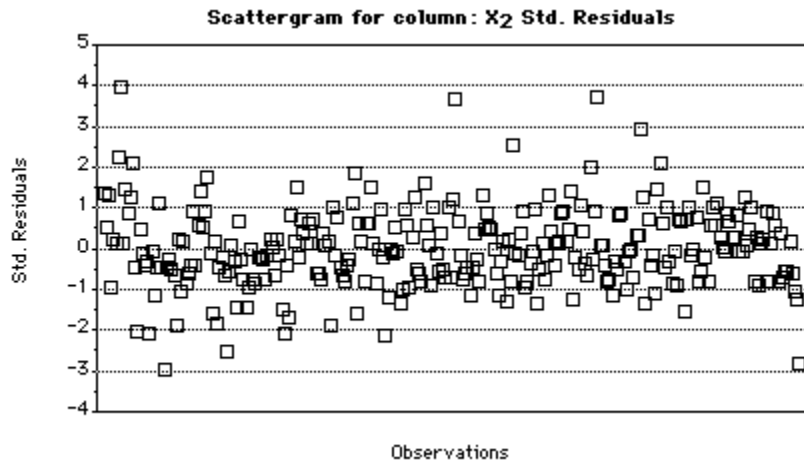


Residual plots, raw and standardized from the census data model.

Raw value residuals



Standardized value residuals

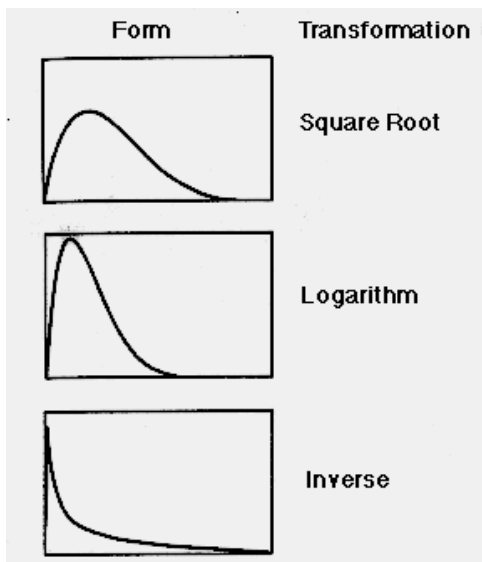


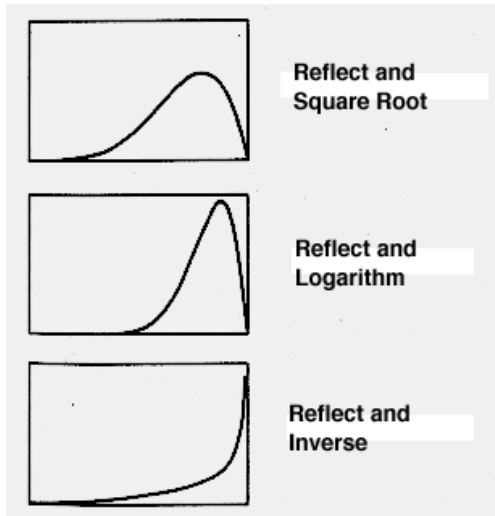
These plots demonstrate that the residuals are from a model where the assumptions were met. The effect of some outliers may need to be assessed.

12. Data Transformation

Transformations are a remedy for outliers, failures of normality, linearity, and homoscedasticity. Yet, caution must still be employed in the usage of transformations due to the increased difficulty of interpretation of transformed variables. The scale of the data influences the utility of transformations, if a scale is arbitrary transformations are more effective, while when the scale is meaningful the difficulty of interpretation increases.

Sample data forms with recommended transformations:





As with many statistical techniques, transformation is an iterative process which requires post calculation evaluation. Check to see that a variable is normally or near-normally distributed after transformation, if not, redo with a more appropriate transformation. Continue attempting transformations until skewness and kurtosis values are nearest zero, or the fewest outliers. The suggested transformations above are intended to bring the distribution closer to normal.

Principal component analysis and factor analysis also serve to transform data in particular situations.

Below is an example of data transformation:

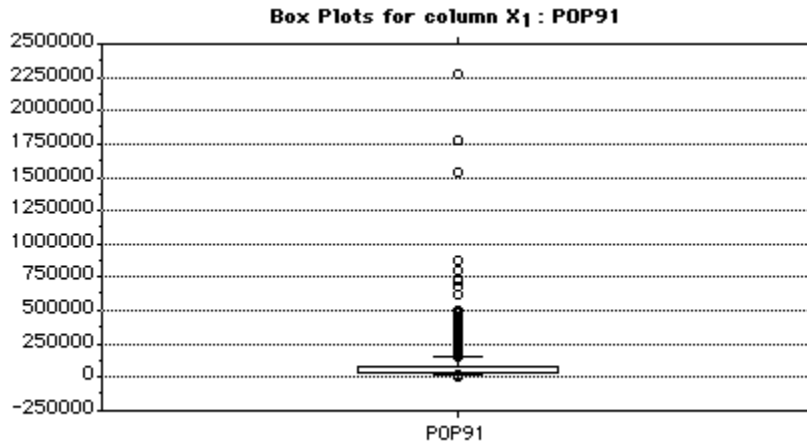
Summary Statistics generated in SYSTAT:

| TOTAL OBSERVATIONS: 290 | | | | | | | | |
|-------------------------|-----|---------|-----------|-----------|-----------|-----------|--------|---------|
| | N | MIN | MAX | MEAN | MEDIAN | STD. DEV | SKEWNS | KURTOS |
| POP91 | 290 | 2153.0 | 2275771.0 | 94127.100 | 37621.0 | 219859.8 | 6.497 | 50.741 |
| NONNAT | 290 | 20.0 | 957370.0 | 14975.310 | 1240.0 | 71315.79 | 9.746 | 112.561 |
| LOWED | 290 | 130.0 | 286205.0 | 10206.034 | 5417.500 | 23951.204 | 9.014 | 92.964 |
| HIGHED | 290 | 135.0 | 336100.0 | 8343.552 | 1710.0 | 29202.842 | 7.650 | 69.487 |
| MUNEMP | 290 | 1.700 | 46.9 | 12.087 | 10.300 | 6.973 | 1.770 | 4.278 |
| MINES | 290 | 0.0 | 33985.0 | 661.810 | 140.0 | 2313.558 | 11.136 | 149.769 |
| GOV | 290 | 130.0 | 94440.0 | 3831.931 | 1350.0 | 9340.04 | 5.954 | 43.370 |
| AVGINC | 290 | 21392.0 | 74142.0 | 44360.403 | 43420.500 | 7215.433 | 0.691 | 1.416 |

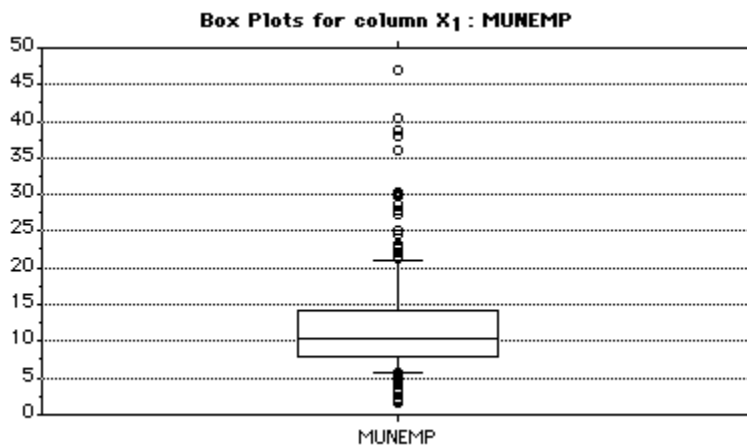
Observe the problems which can be seen in the summary data, such as: scales which vary, ranges which differ, skewness and kurtosis not near zero, means and standard deviations which further reflect the skewed data. Further steps to be taken when viewing the univariate descriptive statistics:

- number of cases all equal, no missing data.
- compare minimum and maximum values to known range.
- assess data for outliers, large range in reference of the mean location and standard deviation, ie. POP91.

Box plot of POP91



as opposed to that of MUNEMP, which is still skewed, yet not as badly.

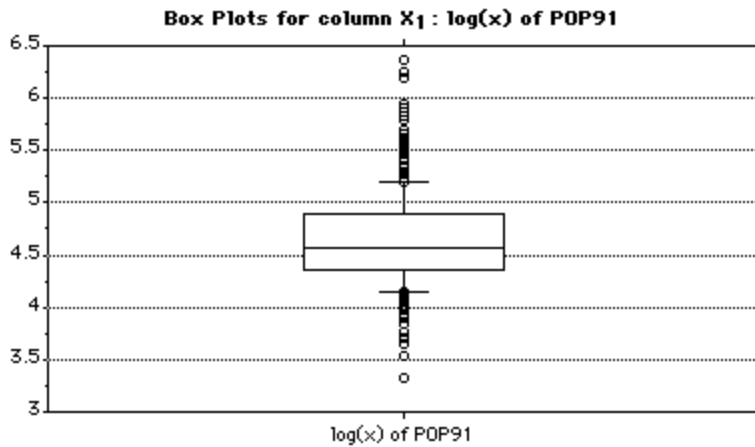


Further to demonstrate the skewed distribution of POP91 is a stem and leaf plot:

STEM AND LEAF PLOT OF VARIABLE: POP91, N = 290

MINIMUM IS: 2153.
LOWER HINGE IS: 21864.
MEDIAN IS: 37621.
UPPER HINGE IS: 78224.
MAXIMUM IS: 2275771.

accordingly POP91 ought to be transformed, consulting the distributions at the beginning of this document which show the different distributions, accordingly, a logarithm transformation will be implemented.

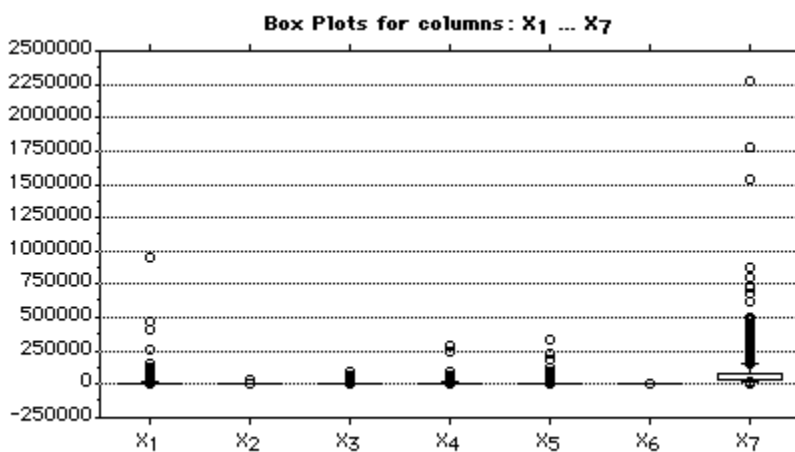


an improved distribution of values is seen with the log transformed population 1991 data.

- observe mean values to see if plausible given the context of the variable, ie AVEINC - average income - approximately \$44360 per year, is a reasonable value.
- get sense of the distribution by comparing mean to median values, the closer together the more symmetrical the distribution.
- *skewness* and *kurtosis*, the closer the distribution is to normal the closer the values of skewness and kurtosis are to zero.

Raw vs. rated values, reduce influence of outliers

RAW



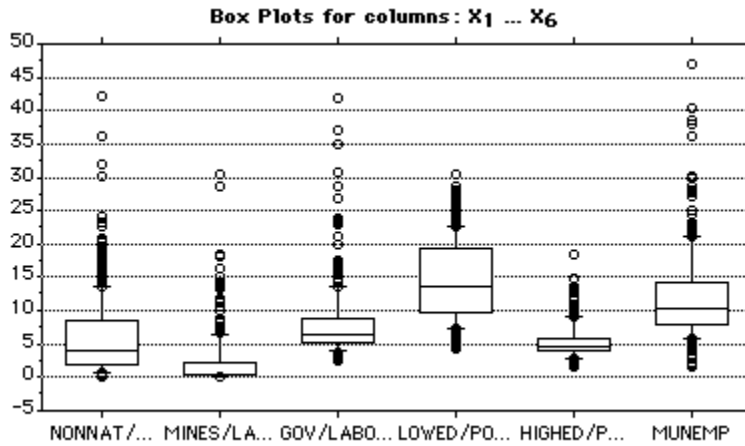
where:

X1 = NONNAT
X2 = MINES

X3 = GOV
 X4 = LOWED
 X5 = HIGHED
 X6 = MUNEMP
 X7 = POP91

The varying ranges for each of the variables reflects the difficulty in comparing of the variables. The larger values of variables such as POP91 will dominate the model.

RATED - why, to reduce the influence of any given variable

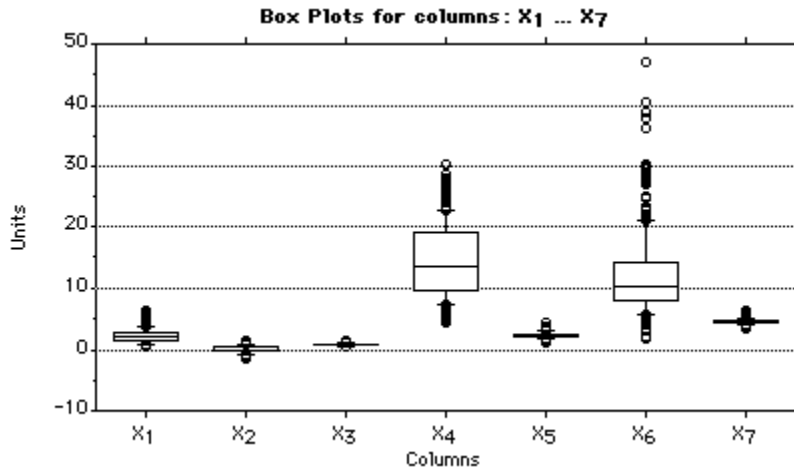


where:

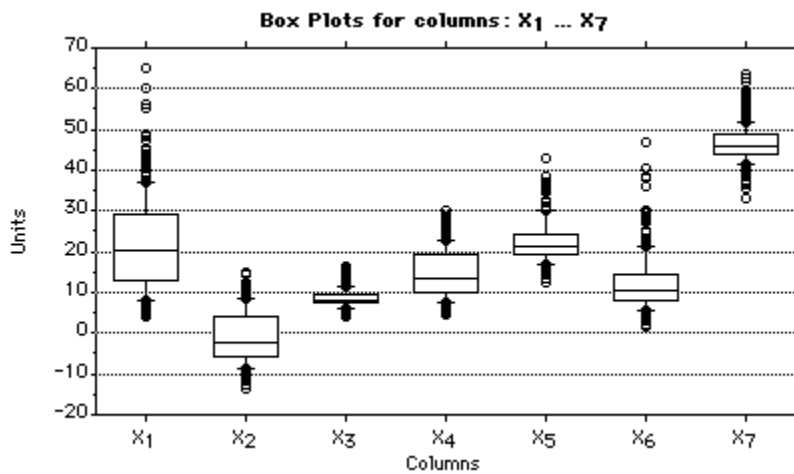
X1 = NONNAT/POP91 (%)
 X2 = MINES/LABOUR (%)
 X3 = GOV/LABOUR (%)
 X4 = LOWED/POP91 (%)
 X5 = HIGHED/POP91 (%)
 X6 = MUNEMP

POP91 will be dealt with in the transformation coming up.

TRANSFORMED



confined data (list) multiplied by 10 to provide that all data within same dynamic range.



where:

- X1 = rated and transformed NONNAT
- X2 = rated and transformed MINES
- X3 = rated and transformed GOV
- X4 = rated LOWED
- X5 = rated and transformed HIGHED
- X6 = raw MUNEMP
- X7 = transformed POP91

14. References

Congalton, 1991; A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data, *Remote Sensing of Environment*, Vol. 37, pp. 35-46.

Davis, J., 1986; *Statistics and Data Analysis in Geology*, (John Wiley & Sons, Toronto, 646p.)

Moore, D., and G. McCabe, 1993; *Introduction to the Practice of Statistics*, (W.H. Freeman and Company, New York, 854p.)

SYSTAT, 1992; *Statistics, Version 5.2 Edition*, (SYSTAT, Inc., Evanston, IL., 724p.)

Tabachnick, B., and L. Fidell, 1989; *Using Multivariate Statistics*, (Harper & Row Publishers, New York, 746p.)

Weslowsky, G., 1976; *Multiple Regression and Analysis of Variance*, (John Wiley & Sons, Toronto, 292p.)

Wetherill, G., 1986; *Regression Analysis with Applications*, (Chapman and Hall, New York, 311p.)

Graduate Geography 616, Suggested Readings List

Clark, W., and P. Hosking, 1986; *Statistical Methods for Geographers*.

Johnston, R., 1978; *Multivariate Statistical Analysis in Geography*.

Haggett, P., *et al.* 1977; *Locational Methods*.

Haining, R., 1990; *Spatial Data Analysis in the Social and Environmental Sciences*

King, L., 1969; *Statistical Analysis in Geography*.

Shaw, G., and D. Wheeler, 1985; *Statistical Techniques in Geographical Analysis* .

Taylor, P., 1977; *Quantitative Methods in Geography*.

Webster, R., and M. Oliver, 1990; *Statistical Methods in Soil and Land Resource Theory*.