A STUDY OF APPLE YIELD RELATIONSHIPS IN 1969 IN THE OKANAGAN AREA OF BRITISH COLUMBIA

by

EWON LEE

B.Sc., Seoul National University, 1964

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the Department
of
Agricultural Economics

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1972

Department of Agricultural Economics

The University of British Columbia
Vancouver 8, Canada

Date September, 1972

ABSTRACT

The purpose of the study is to determine which factors contributed to the production of apples in the Okanagan area during the year 1969.

Regression analysis is used in an attempt to quantify yield relationships. A comparison is made among different tree-size categories in order to determine whether it is necessary to fit separate regression equations instead of using the data for the three groups in a single regression equation. For this purpose an Equality of Slope Test is performed. The outcome of the test shows that there are no significant differences among corresponding coefficients in the equations for tree-size categories. Hence it is feasible to combine them into one equation.

For the regression analysis, two different types of yield relationships are employed: one is a Cobb-Douglas function linear in the logarithms and the other is a quadratic function.

Both functions include a dependent variable, namely, yield per acre and seven independent variables; that is, density, age, value of fertilizer applied, value of spray applied, pruning and thinning labour hours, geographical dummy, and tree-size index. These independent variables are measured on a per-acre basis except in the case of age, geographical dummy and tree-size index.

The data, which consists of cross-section informa-

tion for 1969 represents one hundred and nineteen sample apple plots. It was derived from personal interviews with apple growers.

The quadratic function poses a problem arising from cross-terms in the equation. It was necessary to modify the function in such a manner that the cross-terms included in the regression equation were justified on biological or economic grounds. The regression results for each type of function used in the analysis are discussed and estimates of coefficients and related standard errors shown. It seems desirable that data should be broken down into apple variety groups because different varieties of apple may well have distinct bearing characteristics. Apple trees in the specific plots under study, however, are made up of a mixture of varieties, thus it is extremely difficult to draw a clear map of acreages occupied by each variety. In attempting to obtain variety data, notwithstanding the mixture of varieties in stands, the original data is broken down under certain assumptions. Also in decomposing apple yields into grade constituents similar problems arise.

Despite these difficulties, tests of differences among average yields are made under stated conditions for varietal, tree-size, apple-grade, and regional categories.

These tests reveal that there are no significant differences in average apple yields for varieties, apple grades and regions, but there are significant differences in the case of different tree sizes. The results of these tests are presented in Chapter VI.

The quadratic form of function seems, within the theoretical framework, to be able to represent satisfactorily the apple yield relationship with the selected independent variables. But, in practice, it does not conform well to the empirical situation; it produces a serious multicollinearity problem from the point of view of statistical inference. The Cobb-Douglas function, however, does not cause such a problem. Apart from this, its application brought in almost all the coefficients corresponding to the basic independent variables except for the coefficient of the tree-size index variable. On this evidence, a tentative conclusion was made in favour of the Cobb-Douglas function for the representation of an apple yield relationship in the Okanagan in 1969.

TABLE OF CONTENTS

CHAPTER		PAGE
I.	INTRODUCTION	1
II.	LITERATURE REVIEW OF APPLE BIOLOGY	6
	Factors Influencing Apple Production	6
	Size of Tree	7
	Soil Condition	8
	Frequency of Frost-injury	9
	Unfavourable Condition at Blossom Time	10
	Pruning	12
	Thinning	13
	Spraying	13
	Density	14
III.	LITERATURE REVIEW OF STATISTICS	17
IV.	DATA	32
	Conditions of Sampling	32
	Sampling Method	33
V.	EMPIRICAL RESULTS	45
	Introduction	45
	Results from the Cobb-Douglas Model	54
	Results from the Quadratic Model	55
	Discussion of the Results from Applying Cobb-Douglas and Quadratic Regression Analyses	63
VI.	TESTING FOR THE DIFFERENCE BETWEEN TWO MEANS	65
	Introduction	65
	Outcome of t-test	68
	Discussion of t-test	73

CHAPTER PAGE

VII. SUMMARY AND CONCLUSION 76

Summary 76

Conclusion 78

BIBLIOGRAPHY 81

APPENDIX 85

LIST OF TABLES

TABLE		PAGE
I.	Classification of Rootstock Vigour	86
II.	Variables Used in Models	87
III.	Estimated Simple Linear Regression Equation .	88
IV.	Evaluation of Size-Controlling Effects of Rootstock, Intermediate Framerwork Stock and Strain of Scion Variety on Total Tree Size in Terms of an Index Value	89
V.	Tree-Size Classification of Initial Total Sample	38
VI.	Tree-Size Classification of Sample Enterprises	39
VII.	Input Data	90
VIII.	Correlation Matrix for Simple Linear Regression with One Hundred and Nineteen Pairs of Observations	91
IX.	Covariance Table for Three Tree-Size Groups .	47
X.	Table for S_i in Terms of C_i'	48
XI.	Results of Equality Slope Test for Three Tree-Size Groups	92
XII.	Significant Coefficients at .05 Level for Cobb-Douglas Model	93
XIII.	Correlation Matrix for Cobb-Douglas Model ...	94
XIV.	Significant Coefficients at .05 Level for Quadratic Model	95
XV.	Correlation Matrix for Quadratic Model Involving only Significant Variables	96
XVI.	Observed and Calculated Values of Apple Yields based on Quadratic Model Involving only Significant Variables	97
XVII.	Results from t-tests for Average-Apple-Yield Differences Relating to Tree-Size Groups	69

TABLE		PAGE
XVIII.	Results from t-test for Average-Apple-Yield Differences Between Regions	70
XIX.	Results from t-tests for Average-Apple-Yield Differences Related to Grades	71
XX.	Results from t-test for Average-Apple-Yield Differences Relating to Variety	73

LIST OF FIGURES

FIGURE		PAGE
1.	Value of Regression in Reducing Variation in Y	27
2.	Polynomial Regression as a Special Case of Multiple Regression	53
3.	Range of Values for Possible $\hat{\beta}_1$ Around Origin When X_1 and X_2 are Highly Collinear	60
4.	Differences in Average Apple Yields Among Tree-Size Groups	68
5.	Difference in Average Apple Yields Between Regions (Across All Tree-Size Groups)	70
6.	Differences in Average Apple Yields Among Apple Grades (Across All Tree-Size Groups)	71
7.	Differences in Average Apple Yields Among Apple Varieties (Across All Tree-Size Groups) .	72

ACKNOWLEDGEMENTS

The author wishes to gratefully acknowledge the generous assistance and guidance provided by his thesis supervisor, Dr. M. J. Dorling. In addition, the stimulating intellectual environment provided by the staff and fellow graduate students at the University of British Columbia was extremely motivating and rewarding.

Whole-hearted gratitude is extended to Mr. K. Acton and staff in the Economics Branch, Canada Department of Agriculture, Vancouver, without whose assistance this thesis would not have been possible.

CHAPTER I

INTRODUCTION

Today the tree fruit industry of British Columbia is centred in the Okanagan valley in a narrow one-hundred-mile strip from Vernon to Osoyoos. Approximately 94% of the British Columbia apple crop is produced in the Okanagan and Similkameen valleys. Approximately 4% of the provincial total is produced in the Creston area, with Creston valley being the major orcharding district of that area. The remaining 2% of the crop is produced in scattered pockets ranging from Vancouver Island to the Lower Mainland, Lillooet, Kamloops, Salmon Arm, and Grand Forks areas.

Due to the ability of apple trees to withstand lower winter temperature than other kinds of fruit trees, apples have long been considered the "backbone" of the tree-fruit industry in British Columbia.

A survey conducted by Professor M. J. Dorling, University of British Columbia, indicated that 91% of apple producers in the Okanagan derived all their farming revenue from tree fruit.[1] During the 1960's and early 1970's most British Columbia apple growers continued to engage in orchard renovation: old trees are being replaced by young trees, and in many instances, obsolete varieties and strains

---

[1]M. J. Dorling, The Okanagan Apple Producer --- His Management Attitude and Behaviour, Department of Agricultural Economics, U.B.C., 1968.

are being replaced by more acceptable ones. The commonly accepted tree spacing of 30 feet x 30 feet of former years is giving way to more dense planting. Newly planted orchards with tree spacing of approximately 12 to 16 feet between rows and 5 to 10 feet in the row are becoming commonplace.

The replanting program which British Columbia apple growers have undertaken should help to place the apple industry of the Province in a stronger position so far as the ability to produce competitively is concerned.

There has been an increased emphasis on lower costs of production; earlier fruiting; easier pruning, thinning, and harvesting; and easier spray penetration, all to ensure competitiveness in response to changing market demands. The standardized Malling vigour-controlling root-stocks seem likely contributors to achieving some of these goals. The need for consideration of these matters led to the initiation of this study --- the purpose of which is to estimate yield relationships with special reference to density and other production influences, by means of regression analysis.

As a preliminary step, various simple linear regression analyses were attempted; discussion of these analyses centre on the apple yield performance in relation to the basic independent factors of production, namely, density per acre, age of trees, the amount of fertilizer, the amount of spray, pruning and thinning labour hours.

The empirical results of these analyses are discussed in Chapter V and the results are shown in Table III in the Appendix. Numerous empirical and logical criteria have been used in the selection of a quadratic function in the study: before this was accomplished, different algebraic models were employed in representing observational data. Selection of a particular form of function was based mainly on two considerations: significance of structural coefficients and best fit.

The best fit was indicated by the magnitude of the coefficient of determination, R^2, assuming that the condition of normally and independently distributed errors was not violated. The logical reasons for selection of the quadratic model are the following: (1) it allows both declining and negative marginal productivity (these conditions are very important from the apple study's point of view because apple yield is assumed to be subject to the law of diminishing returns); (2) it does not impose such strict restraints on a yield relationship as the Cobb-Douglas and Spillman equations; (3) a maximum total yield is defined.[2]

In summary, deciding both the functional form, and which variables to omit and which to retain was done on the basis of the logic, including consideration of physical and biological relations and statistical probability levels. If the R^2 is satisfactory and the logic of the production

---

[2]See E. O. Heady and J. Dillon, Agricultural Production Functions, pp. 75-78, Ames, Iowa: Iowa State University Press, 1961.

situation does not dictate that the excluded variable must be included, the new regression estimates may be regarded as serving satisfactorily. The quadratic function was chosen by the above criteria.

More often than not, selection among algebraic forms of equations is no less difficult than decision with respect to the significance level at which variables will be omitted from the quadratic models being examined.

This study was designed to estimate yield equations and identify the most important contributing factors in apple production in the Okanagan area of British Columbia in 1969.

An assumption was made that all independent variables were measured without errors and the dependent variable was a stochastic variable exhibiting observed disturbance. Sampling methods and methods of derivation of data are outlined in Chapter IV. Chapter V is devoted to a preliminary review of two different forms of function in terms of their relevancy to the study. The Equality of Slope Test is also reviewed to ensure that there can be justification for combining the three different tree-size equations into a single equation. The purpose of the Equality of Slope Test is in relation to K linear regression equations in m independent variables: $Y_i = b_0 + b_1^i x_1 + ---- + b_m^i x_m$, $i = 1, 2, ----, k$.[3] It tests the hypothesis $H_0: b_j^1 = b_j^2 =$

[3]As is explained in Chapter III regression equations are expressed in the early thesis chapters as having small 'x' (deviation) variables. These forms are convenient conceptualizations.

---- = b_j^k, j = 1, 2, ----, m. A regression equation $Y = b_0 + b_1x_1 + ---- + b_mx_m$ is found for each of k groups of sample units and an F-test is carried out to determine whether differences in the estimated coefficients among groups is due to sampling errors or real differences.

The results of the analysis involving a selected regression equation and related discussion are presented at the end of Chapter V. In Chapter VI, numerous tests are made concerning differences among average apple yields with regard to different tree size, apple varieties, apple grades, and regions as important sources of influence. Finally, Chapter VII presents a summary of the main conclusions and implications of the study.

CHAPTER II

LITERATURE REVIEW OF APPLE BIOLOGY

Factors Influencing Apple Production

There tends in practice to be two main aspect of an apple enterprise which focus interest: one is biological, and the other is economic. The primary purpose of this study was to outline some of the most important factors associated with the yield performance of specific apple enterprise plots. There are many factors influencing apple production, perhaps too many to pinpoint them all. A high level of management in operating apple orchards may well be conducive to increasing the level of production. The same can be said of size of operation, type of machinery available, and so forth. But these factors are difficult to quantify and this makes fitting a regression equation in which they are represented infeasible. On account of this difficulty, attention will be confined only to the quantifiable factors of production. It goes without saying that this procedure cannot be immune from a danger of oversimplification.

In any study of orchard production, the following variables, among others, are important:[1] tree size, soil

---

[1]See J. C. Folger and S. M. Thomson, The Commercial Apple Industry of North America, ed. L. H. Baily, pp. 339-347. New York: Macmillan Co., 1921.

See R. Bush, Tree Fruit Growing, revised by E. G. Gilbert. Prepared in conjunction and collaboration with the Royal Horticultural Society. Penguin Books, 1962.

See D. W. Ware, E. D. Woodward and H. W. Trevor, A Study of Apple Production in the Okanagan Valley of British Columbia, Canada Department of Agriculture, Marketing Service - Economic Division Ottawa, January 1952.

conditions, the frequency of frost-injury, unfavorable conditions at blossom time, pruning, thinning, spraying and density. There are discussed in turn.

Size of the Tree

Apple varieties are propagated by means of grafting or budding. Therefore it is necessary to have rootstocks on which to graft or bud scion-wood of selected varieties.

Most apple trees consist of two distinct parts, rootstock and scion variety. In some instances, particularly with the more tender apple varieties, it may be desirable to use trees with a winter-hardy trunk and/or framework. Wood of a winter-hardy variety is used for that purpose and if it differs from the rootstock, it is referred to as an intermediate stock. Consequently, trees with an intermediate stock contain three distinct sections: rootstock, intermediate stock and scion variety. Rootstocks are given first consideration, since size controlling roots provide the most practical means of determining ultimate tree size.

The introduction of dwarfing rootstocks and spur-type varieties showing compacter growth provided the opportunity to adopt new orchard planting systems. Classification of rootstock vigour is presented in Table I in the Appendix. Semi-standard, semi-dwarf, and dwarf trees make it possible to combine high tree population per acre with early and high yield. Tukey, Extension Horticulturist at Washington State University, has shown that the smaller tree allowing larger numbers of trees per acre has a greater potential for high early yield.[2]

Fisher alludes to similar facts in his statement: "Many old, ailing, out-of-date orchards required renovation. In replanting these blocks, and also in bringing in new land, the grower has become increasingly conscious of the need for smaller easier to handle trees, more attractive to labour, and capable of producing high early fruit returns."[3]

Brase and Way have found that small apple trees, because of reduced bearing area, will produce less fruit per tree than large standard trees, but as more trees can be planted, larger or at least as large yields per acre will be produced.[4]

Soil Condition

The prerequisite of an orchard soil is that it be well-drained. Soils are the products of the environmental conditions under which they have developed. These conditions involve mineral materials as well as topographic, climatic and biological phenomena. Well-drained soils which reflect

---

[2] R. B. Tukey, "Implications of Economics on Orchard Management", The 1969 Apple Forum, Published Proceedings of the First British Columbia Fruit Growers' Association-sponsored Horticultural Conference. pp. 59-60 (November 1969).

[3] D. V. Fisher, High-Density Orchards for British Columbia Conditions, Research Station Summerland, British Columbia Research Branch, Canada Department of Agriculture, March 1966.

[4] K. D. Brase and R. D. Way, Rootstocks and Methods used for Dwarfing Fruit Trees, New York State Agricultural Experiment Station, p. 783, 1959.

the forces of soil genesis, climate and vegetation, are classified as zonal soils. The zonal distinction is believed to be due to a variable moisture and temperature relationship characterizing mountainous country. The soil of the lower-part slope in the Okanagan belongs to the Glenmore clay-loam formation; the soil of the upper part belongs to the Oyama loamy-sand formation. Both are classified as dark brown soils by Kelly and Spilsbury.[5]

The apple tree in commercial production also requires a number of mineral elements, e.g., magnesium, potassium, manganese, clacium, sulphur, iron, boron, copper, and zinc. These elements are frequently applied in the form of fertilizer and spray compounds.

The Frequency of Frost-injury

In considering geographic and climatic factors, mention should be made of the importance of frost-injury in orchards in certain locations. Orchards in most Okanagan areas receive occasional damage from frost. However, some areas are more susceptible than others, and for the region as a whole, the micro-climate is quite variable. Ware established a table indicating different frost-free periods corresponding to areas in the Okanagan as follows:[6]

---

[5] C. C. Kelly and R. H. Spilsbury, "Soil Survey of the Okanagan and Similkameen Valley of B.C.", Report No. 3 of B.C. Survey. The British Columbia Department of Agriculture in cooperation with Experimental Farm Service, Dominion Department of Agriculture. pp. 20-71, 1949.

[6] D. W. Ware, Organization and Returns of Stone Fruit and Pear Enterprises in the Okanagan Valley, B.C. 1949-1950, Department of Agriculture Economic Division, Marketing Service, Ottawa. 1952. pp. 5-6.

Area	Frost-free Period in Days
Kelowna	150
Summerland	176
Penticton	152
Oliver	162
Keremeos	188

The frost-free period as defined for the above data is the number of days between the last date in the spring on which the temperature of 32° F. was recorded and the first similar condition in the fall of the same year. The figures given are averages obtained for a ten-year period.

Unfavourable Conditions at Blossom Time

It is well known that variations in climate account for considerable variations in crop yields. Temperature is of extreme importance at all seasons of the year in the growing of apples. Winter temperature may be so low as to result in injury to the buds or the wood of the tree. On the other hand, some cold winter temperature is required to ensure vernalization so that trees leaf out normally in the spring. Temperature in the spring may also be a critical factor. Temperatures of 26° or 27° F. for periods as short as an hour or so can cause damage to flowers. Thus, climate is a most important variable which either implicitly or explicitly enters any supply equation for an agricultural crop.[7]

It is difficult to find an appropriate index for measuring the influence of climate on apple production. The most convenient measure might well be temperature if that were summarized in a convenient form. For this study a geographical dummy variable was introduced into the analysis which was intended to represent the influence of weather.

The correlation existing between rainfall and size of apple crop in Nova Scotia has been found to be negative by Longley --- limits of the population correlation coefficient for the seven-year period 1913-1929 inclusive, involving the May to October period, were estimated as -0.572 ± 0.110.[8] Hence, decreasing rainfall in the summer months tended to be associated with an increasing crop.

It is of interest to note that during the months in which spraying and dusting operations are done, the hours of sunshine are a critical factor in the production of apples. A combination of more hours of sunshine and less rainfall during the months of May, June and July results in a more effective production of tree foliage and better control of insects and diseases.

---

[7] J. P. Doll, "An Analytical Technique for Estimating a Weather Index from Meteorological Measurements", Journal of Farm Economics, Vol. 49, No. 1, February 1967.

H. S. Lawrence, "The Effect of Weather on Agricultural Output": A Look at Methodology, Journal of Farm Economics, Vol. 46, No. 1, February 1964.

A. Koutsogianne-Kokkova, An Econometric Study of the Leaf Tobacco Market of Greece, pp. 164-166, Athens, 1962.

[8] W. V. Longley, Some Economic Aspects of the Apple Industry in Nova Scotia. A Thesis for the Degree of Doctor of Philosophy, pp. 22-23, Nova Scotia Department of Agriculture Bulletin No. 113, 1932.

Pruning

A good tree framework of desired size, form and strength is necessary to obtain the maximum number of well spaced branches and spurs in as small an area as possible and still allow the fruit plenty of space to grow. Pruning admits light and air, allows easier spraying and picking, and thus improves fruit buds. Pruning is essential to avoid a bare-wood condition and to induce fruiting near the trunk or main branches. Correct pruning helps to make possible the rigid cordon shape or the permanently dwarfed pyramid tree, but it must be employed in the right way on the right variety of tree, planted in suitable soil, if the best results are to be expected. It is no use expecting all varieties of tree to conform to the same standard; one must adapt one's pruning to take advantage of the natural habit of the particular variety.[9]

Apple trees may be said to pass through three distinct periods: (1) formative period, (2) transitive period, and (3) fruiting period.[10] Appropriate pruning treatment changes materially with each of these periods. It is during the formative period that the tree devotes its energies to wood growth. The proper selection, distribution and training of branches during this time determines the ability of the tree to bear heavy crops of fruit in later years. All pruning during the transitional period is to

---

[9]Bush, op. cit., pp. 115-116.

[10]Folger, et al., op. cit., p. 283.

develop and maintain a liberal supply of fruiting wood, well distributed throughout the entire tree.

Thinning

The thinning of apples is no more than a form of pruning. If all the fruit on an apple tree showing a heavy fruit set were allowed to mature, small misshapen fruits and limb breakage would result. This is because fruit bud formation takes place early in the season during a period of extreme competition between fruit buds and young fruit for available food supplies. Early removal of surplus fruits removes much of this competition.[11]

Spraying

Orchards must be sprayed regularly and thoroughly in order to protect the fruit from serious insect or disease damage. Depending on the nature and extent of the infestations, apples require from four to seven sprays a year.

Among entomologists, however, there are two schools of thought. The one believes that wholesale liquidation by poisonous sprays is desirable. In contrast, the other school regrets the massacre of many beneficial insects and hopes that biological control will prove superior.

There no doubt exists a danger of inducing immune races of insects; today, there are several pests which, having in the past been exposed to certain sprays, have now

---

[11] Ibid., op. cit., p. 283.

developed a degree of immunity. However, apple growers cannot expect their particular orchards to be free from attacks of insects and diseases which occur elsewhere, and any who omit spraying are unlikely to produce marketable fruit.

Density

Van Roechoudt has written: "At the end of the sixth growing season, on an acre basis, there was a wide variation in yields from each of the different planting concepts. The trees planted as hedgerows on M. VII rootstock at the density used had produced 25.9 times more fruit. The yield was related to the planting concept, the number of trees per acre, the system of pruning and training followed and the type of rootstock used."[12]

Harris and Woods have reported from their investigations at the Canada Department of Agriculture Experimental Farm, Saanichton, B.C., that apple trees at higher density on M. IX rootstock grow well, produce heavily with high quantity fruit at an age when standard trees are far from being in a state of commercial production.[13]

Intensive planting of apple trees implying high density per acre will involve a high investment cost. Of primary consideration, however, is the ability of the crop

---

[12] L. L. Van Roechoudt, Some Factors Which Influence the Use of Dwarf and Semi-Dwarf Apple Trees for Commercial Orchards in the Okanagan Valley of B.C. Unpublished Master's Thesis, The University of British Columbia, 1962.

[13] J. H. Harris and J. J. Woods, Dwarf Apple Trees on Vancouver Island, Experimental Farm Research Branch, Saanichton, B.C., 1958.

to return a profit on the investment. Smaller trees inherently produce fruit at an earlier age; the larger number of trees per acre can result in a significantly higher yield per acre. Tukey states: "One of the most positive methods of increasing yield in the early years of an orchard is to increase the number of trees per acre, and increasing the tree population may be one of the most effective means of counteracting the problem of obsolescence and replanting old orchard sites."[14]

[14] Tukey, op. cit., p. 58.

CHAPTER III

LITERATURE REVIEW OF STATISTICS

Modern statistics are based upon probability. There are a number of conflicting ideas about this concept, which is fundamental for scientific methodology.

Some authors hold that probability statements refer to a proposition and are hence logical and not empirical. This concept refers to our rational degree of belief in a theory or hypothesis on the basis of empirical evidence.

Keynes, for example, expounds in his treatise on probability as follows: "What we know and what probability we can attribute to our rational beliefs is, therefore, subjective in the sense of being relative to the individual. But given the body of premise which our subjective powers and circumstances supply to us, and given the kinds of logical relations upon which arguments can be based and which we have the capacity to perceive, the conclusion,which it is rational for us to draw, stands to these premises in an objective and wholly logical relation. Our logic is concerned with drawing conclusions by a series of steps of certain specified kinds from a limited body of premises."[1]

Another and entirely different probability concept refers to the relative frequency of an event, as the

---

[1]J. M. Keynes: A Treatise on Probability, p. 18.

number of trials increases indefinitely. The econometrician may, for instance, consider the relative frequency of business failures --- that is, the percentage of businesses which fail each year. He may talk about the probability of a business failure as the limit of the relative frequency of failures as the sample becomes larger and larger. Since the first probability concept is not yet useful for any but the simplest problems of statistical inference, only the second concept is relevant in the case of statistical tests in the study.

The fundamental purpose of regression analysis is to estimate the relationship between the dependent and independent variables. Once the relationship between these variables has been quantitatively estimated, we may wish to know the goodness of fit of the relationship.

It is impossible to estimate the relationship between the variables without first making some assumptions or deductions about the form of the relationship. To illustrate, consider a simple linear regression equation, $Y_i = a + bx_i$, where $i=1,2,----n$ and where $x_i=(X-\overline{X})$.[2] One advantage of measuring X_i as deviations from their mean is that the mathematics will be simplified because the sum of the new x values equals zero --- that is $\Sigma x_i=0$. This will become convenient later on in the proof of $E(b) = \beta$, $var(b) = \sigma_y^2/\Sigma x_i^2$. Also, in the process of inverting $\mathbf{X'X}$, where $\mathbf{X}$ is a matrix consisting of all x observation and $\mathbf{X'}$ is a

[2] See R. J. Wonnacott and T. M. Wonnacott, <u>Econometrics</u>, pp. 245-246. John Wiley and Sons, Inc., 1970.

matrix consisting of all x observations and $\mathbf{X}'$ the transpose of $\mathbf{X}$, the measurement of x values in deviation form is shown to be very convenient. Suppose that an experiment could be repeated many times at a fixed value of x. Then there would be observed some statistical fluctuation of the Y values clustered about a central value forming a sub-population. The probability function of Y for a given x, we shall call P(Y/x). There will be a similar probability function for Y at any other experimental level of x. Consequently, probability functions for Y_i at the various levels of x_i will be $P(Y_i/x_i)$.

To keep the problem manageable, let there be a reasonable set of assumptions about the regularity of these sub-populations. These assumptions may be written concisely as follows: the random variables Y_i are statistically independent, with mean $\alpha+\beta x_i$ and variance σ_y^2. On occasion, it is useful to describe the deviation of Y_i from its expected value as the error or disturbance term U_i, where the U_i are independent random variables, with mean 0 and variance σ_y^2. No assumption is yet made about the shape of the distribution of U_i provided it has a finite variance. The error term may be regarded as the sum of two components:

1. Measurement Error.

In measuring crop yield, there may be an error resulting from careless harvesting or inaccurate weighing.

2. Stochastic Error.

Disregarding measurement error, there would still be some unpredictable differences in yields, for example, in an experiment using the same rate of fertilizer application. Assume that the situation is such that there are no large measurement errors in the variables. However, there are certain variables which ought to appear in the equation but have been left out. Omission of the latter results in rather large errors in the equations.[3]

If the entire populations of values (x_i, Y_i) are known, it is possible to compute the exact values of the regression parameters α, β and σ_y^2. Determination of least squares is the most acceptable method for fitting a straight line. The method of least squares requires that the estimators (a,b) be selected in such a way that the sum of the squared deviations of $\hat{Y}_i = a + bX_i$ from the fitted regression line be a minimum --- that is minimize $e_i^2 = (Y_i - a - bx_i)^2$, where e is the error term. For testing hypotheses, it will be necessary to know how the estimators a and b are distributed around their parameters, α and β. The least squares estimators a and b are then the best linear unbiased estimators of α and β. That is, to sum up:

$$E(a) = \alpha$$

$$\mathrm{Var}\ (a) = \sigma_y^2/n$$

$$E(b) = \beta$$

$$\mathrm{Var}\ (b) = \sigma_y^2/\Sigma x_i^2$$

[3] T. Haavelmo, "The Probability Approach in Econometrics", Econometrica, Vol. 12, 1944, Supplement.

H. B. Mann and A. Wald, "On the Statistical Treatment of Linear Stochastic Difference Equations", Econometrica, Vol. 11, p. 173, 1943.

where E and Var stand for expected value and variance respectively. These properties have been proved with the use of Gauss-Markov Theorem without making any assumption about the shape of the distribution of the error term.[4] Since the slope coefficient b is usually of more interest to us than the intercept coefficient a, we shall concentrate on the slope. Proof of $E(b) = \beta$ and $Var(b) = \sigma_y^2/\Sigma x_i^2$ alone is as follows. The formula for b may be rewritten as:

$$b = \Sigma(x_i/K)Y_i \tag{3-1}$$

where

$$K = \Sigma x_i^2 \quad . \tag{3-2}$$

Thus,

$$b = \Sigma w_i Y_i = w_1 Y_1 + w_2 Y_2 + \cdots w_n Y_n \tag{3-3}$$

where

$$w_i = x_i/K \tag{3-4}$$

From the theory of linear transformations, it follows that:

$$E(b) = w_1 E(Y_1) + w_2 E(Y_2) + \cdots w_n E(Y_n) = \Sigma w_i E(Y_i) \tag{3-5}$$

Noting that the variables Y_i are assumed independent, it follows that

$$Var(b) = w_1^2 Var\ Y_1 + w_2^2 Var\ Y_2 + \cdots w_n^2\ Var\ Y_n = \Sigma w_i^2 Var\ Y_i \tag{3-6}$$

Using the mean from (3-5) and $E(Y_i) = \alpha + \beta x_i$ as assumed previously, then

$$E(b) = \Sigma w_i(\alpha + \beta x_i) = \alpha \Sigma w_i + \beta \Sigma w_i x_i$$

and noting equation (3-4), then

[4] Wonnacott and Wonnacott, op. cit., pp. 48-51.

$$E(b) = (\alpha/k)\Sigma x_i + (\beta/k)\Sigma(x_i)x_i \ .$$

But, since Σx_i is zero, then

$$E(b) = 0 + (\beta/k)\Sigma x_i^2 \ .$$

Furthermore, from equation (3-2)

$$E(b) = \beta$$

From equation (3-6) and from $Var(Y_i)=\sigma_y^2$ as assumed previously,

$$Var(b) = \Sigma w_i^2\sigma_y^2 = \Sigma(x_i^2/k^2)\sigma_y^2 = (\sigma_y^2/k^2)\Sigma x_i^2$$

Again, noting equation (3-2),

$$Var(b) = \sigma_y^2/\Sigma x_i^2 \ .$$

Recalling the assumption that Y_i values are statistically independent and also that b is a linear combination of all Y_i (that is $b = \Sigma x_i Y_i/\Sigma x_i^2$), it follows that the shape of the b distribution will also be normal. The normality assumption of the error term is required only for small sample estimations. Without assuming that the Y_i are normally distributed, as sample size increases, the distribution of b will usually approach normality, this can be justified by a generalized form of the Central Limit Theorem. If we have specified the form of the distribution of the error terms in our regression model, then the method of least squares is justified by the method of maximum likelihood (which could also have been used to obtain estimators α and β).

For generality, suppose that we have a sample of size n. We wish to know:

$$P(Y_1, Y_2 ---- Y_n) \qquad (3\text{-}7)$$

That is, we wish to know the likelihood or probability density of the sample we observed, expressed as a function of the possible population values of α, β and σ_y^2 . Therefore, first consider the probability density of the first value of Y_i which is

$$P(Y_1) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(\frac{1}{2}\sigma_y^2)(Y_1-(\alpha+\beta x_1))^2}, \qquad (3\text{-}8)$$

where $e = 2.71828$

This is simply the normal distribution of Y_1, with its mean $(\alpha+\beta x_1)$ and variance (σ_y^2) substituted into the appropriate positions. The independence of the Y_i values justifies multiplying all these probability densities together to find the joint probability density:

$$P(Y_1, Y_2, ----Y_n) = \qquad (3\text{-}9)$$

$$\left(\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(\frac{1}{2}\sigma_y^2)(Y_1-(\alpha+\beta x_1))^2}\right)\left(\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(\frac{1}{2}\sigma_y^2)(Y_2-(\alpha+\beta x_2))^2}\right)$$

$$---- = \prod_{i=1}^{n}\left(\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(\frac{1}{2}\sigma_y^2)(Y_i-(\alpha+\beta x_i))^2}\right)$$

where $\prod$ represents the product of n factors. Using the familiar rule for exponentials, the product of equation (3-9) can be expressed as follows:

$$P(Y\ , Y\ , ----Y_n) = \left(\frac{1}{2\pi\sigma_y^2}\right)^{n/2} e^{-(\frac{1}{2}\sigma_y^2)(Y_i-(\alpha+\beta x_i))^2} \qquad (3\text{-}10)$$

Recalling that with the observed Y_i speculation is made concerning the values of α, β and σ_y^2 , then, to emphasize this, the equation (3-10) is renamed the likelihood function:

$$L(\alpha,\beta,\sigma_y^2) = \left(\frac{1}{2\pi\sigma_y^2}\right)^{n/2} e^{-(\frac{1}{2}\sigma_y^2)(Y_i-\alpha-\beta x_i)^2} \qquad (3\text{-}11)$$

Therefore, the question is: Which values of α and β make L largest? The only place that α and β appear is in the exponent. Moreover, maximizing a function with a negative exponent involves minimizing the algebraic magnitude of the exponent. Designating our estimators as a and b, the problem is to select values for these that minimize

$$(Y_i - a - bx_i)^2 \quad . \tag{3-12}$$

The conclusion that follows is that maximum likelihood estimates are identical to least squares estimates when the regression model has a normally distributed error.

So far the independent variable x has assumed a given set of fixed values. However, in many cases, x cannot be controlled in this manner. Thus if we are examining the effect of rainfall on yield, it must be recognized that x (i.e., rainfall) is a random variable, completely outside our control. The method of least squares is still valid whether x is a fixed or a random variable, provided that we assume that the distribution of x does not depend on α, β, and σ_y^2 , and that the error terms are normally distributed and independent of the x's (3-13)

Of these assumptions, we must emphasize the independence of x and U. It can be shown that the maximum likelihood and least squares estimates coincide and may be applied regardless of whether the independent variable x is fixed or random, provided x is independent of the error and parameters in the equation being estimated. The likelihood of our sample now involves the probability of observing both x and Y. Therefore, if the x_i are independent, the likelihood function is

$$L = P(x_1)P(Y_1/x_1)P(x_2)P(Y_2/x_2) \qquad (3\text{-}14)$$

Since the error terms are considered normal,

$$L = P(x_1) \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(\frac{1}{2}\sigma_y^2)(Y_1-\alpha-\beta x_1)^2} P(x_2) \frac{1}{\sqrt{2\pi\sigma_y^2}}$$

$$e^{-(\frac{1}{2}\sigma_y^2)(Y_2-\alpha-\beta x_2)^2} \qquad (3\text{-}15)$$

Collecting the exponents,

$$L = P(x_1)\text{---}\left(\frac{1}{2\pi\sigma_y^2}\right)^{n/2} e^{-(\frac{1}{2}\sigma_y^2)\Sigma(Y_i-\alpha-\beta x_i)^2} \qquad (3\text{-}16)$$

Since according to equation (3-13), P(x) does not depend on the parameters α, β, and σ_y^2 , the problem of maximizing this likelihood function reduces to the minimization of the exponent in equation (3-11).

It is of interest to note what would happen if the independent variable x is correlated with the error terms. Reconsider the model,

$$Y = \alpha + \beta x + U \qquad (3\text{-}17)$$

By taking the covariances of x with each of the variables in the equation, the following results,[5]

$$S_{xy} = S_{xx} + S_{xu} \qquad (3\text{-}18)$$

In order to estimate β, S_{xy} is divided by S_{xx} (variance of

[5] Wonnacott and Wonnacott, op. cit., pp. 149-157.

x) such that

$$S_{xy}/S_{xx} = b + S_{xu}/S_{xx} \tag{3-19}$$

From the observation of x and Y, S_{xy} and S_{xx} are easily calculated. However, U is unobservable, so that S_{xu} cannot be evaluated. Therefore, if we can assume that S_{xu} is small enough to neglect, we will obtain the estimator

$$S_{xy}/S_{xx} = b \tag{3-20}$$

We recognize this as the least squares estimator.[6] That is, from equation (3-20), the least squares estimator is justified under conditions that $S_{xu} \overset{P}{\rightarrow} 0$ while $S_{xx} \overset{P}{\rightarrow}$ nonzero (where $\overset{P}{\rightarrow}$ is defined as approaches in probability as $n \rightarrow \infty$).

We have so far dealt with regression analysis relevant to this study. But, interest may also focus on correlation analysis --- that is the degree to which variables are related or associated. Simple correlation analysis yields only one coefficient --- and index number --- designed to give an immediate picture of how closely two variables move together. In correlation analysis, cause and effect relations are unimportant.

A distinction between regression analysis and correlation analysis must be made to avoid confusion which may arise from the subtlety of the propositions involved in both analyses. In regression analysis, all the independent variables are assumed fixed. They do not occur in a probabilistic way. On the other hand, correlation analysis is

---

[6] As noted before $b = \Sigma Yx/\Sigma x^2 = \Sigma yx/\Sigma xx = \sqrt{\Sigma yx/n-1} \,/\, \sqrt{\Sigma xx/n-1} = S_{yx}/S_{xx}$.

concerned mainly with random variables. Independent variables must have a respective probability distribution. In view of these differences, r^2 values can be adequately interpreted only in a correlation analysis. Yet, since correlation and regression are so closely related mathematically, correlation often becomes a useful aid in regression analysis. Specifically, consider the relation between the estimated correlation coefficient r, and the estimated regression slope coefficient b. It was shown that

$$b = \Sigma xy / \Sigma x^2 \qquad (3\text{-}21)$$

Noting that both x and y are defined as deviations, then

$$r = \Sigma xy / \ \ \Sigma x^2 \Sigma y^2 \qquad (3\text{-}22)$$

Then

$$b/r = \sqrt{\Sigma x^2}\sqrt{\Sigma y^2} \Big/ \Sigma x^2 = \sqrt{\Sigma y^2 / \Sigma x^2} \qquad (3\text{-}23)$$

If we now divide both the numerator and denominator inside the square root sign by (n-1),

$$b/r = \sqrt{(\Sigma y^2/n-1)/\Sigma x^2/n-1)} = S_y/S_x \qquad (3\text{-}24)$$

or

$$b = r(S_y/S_x) \qquad (3\text{-}25)$$

This close correspondence between b and r will be of utmost importance in the subsequent argument as to which tool is the more powerful --- regression or correlation analysis.

Consider fitting a regression line to the scatter of observations (x_i, Y_i). This is represented in Figure 1, where $\hat{Y}_i$ = the regression estimate of Y_i .

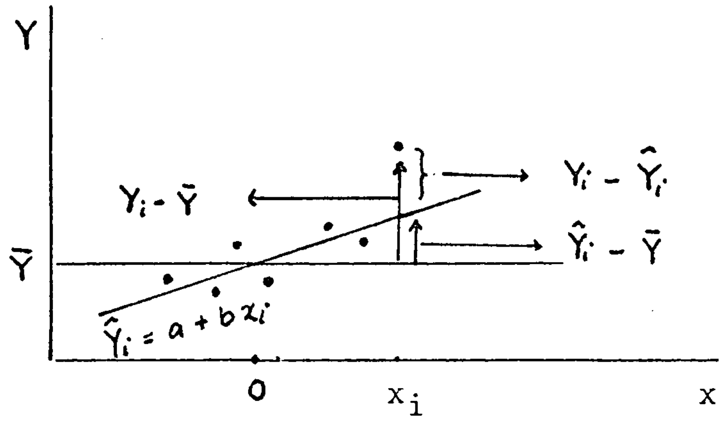

Figure 1.
The value of regression in reducing variation in Y.

Now, the best prediction of a Y without knowing x would be the average observed value ($\bar{Y}$). At x_i, it is clear from this diagram that we would make a very large error --- namely $(Y_i - \bar{Y})$ --- the deviation of Y_i from its mean. However, once the regression equation has been calculated, we predict Y to be $\hat{Y}_i$ and this reduces the error, sinze $(\hat{Y}_i - \bar{Y})$ which is a large part of the deviation has now been "explained". Therefore, this leaves only a relatively small "unexplained" deviation $(Y_i - \hat{Y}_i)$. Total deviation of Y is the sum:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i), \text{ for any } i \qquad (3\text{-}26)$$

It follows that

$$\Sigma(Y_i - \bar{Y})^2 = \Sigma(\hat{Y}_i - \bar{Y})^2 + \Sigma(Y_i - \hat{Y}_i)^2 \qquad (3\text{-}27)$$

where variation is defined as the sum of the squared deviations. Since $(\hat{Y}_i - \bar{Y}) = \hat{y}_i = bx_i$, it is convenient to rewrite equation (3-27) as

$$\Sigma(Y_i - \bar{Y})^2 = b^2\Sigma x_i^2 + \Sigma(Y_i - \hat{Y}_i)^2 \qquad (3\text{-}28)$$

The fact that explained variation is the variation accounted for by the estimated regression coefficient b is now clarified by the above equation. The procedure of decomposing total variation and the analysis of its components is called "analysis of variance applied to regression". From the foregoing, a null hypothesis test on β may be constructed. The question is then, whether the ratio of the explained variance to unexplained variance is sufficiently large to reject the hypothesis that Y is unrelated to x. Specifically, a test of the hypothesis $H_0:\beta = 0$ involves forming the ratio "F" equal to variance explained by regression divided by the unexplained variance equal to:

$$b^2(\Sigma x_i^2/S^2) \qquad (3\text{-}29)$$

where S^2 is the sample variance of Y. It must be emphasized that this is just an alternative way of testing the null hypothesis with the use of the "t-distribution":

$$\text{calculated "t"} = b/\sqrt{S^2/\Sigma x_i^2} \qquad (3\text{-}30)$$

For the "t-distribution" to be strictly valid, the strong assumption is made that the distribution of Y_i is normal. Note that the "F" and "t" distributions are related, generally, as follows: $F = t^2$, where there is one degree of freedom in the numerator of F. The variation in Y will now be related to r. It follows from equation (3-25) that

$$b = r\sqrt{\Sigma Y_i^2/\Sigma x_i^2} \quad .$$

Then, substituting this value for b in equation (3-28)

$$\Sigma(Y_i - \overline{Y})^2 = r^2 \Sigma y_i^2 + \Sigma(Y_i - \hat{Y}_i)^2 \qquad (3\text{-}31)$$

Noting that y_i^2 is by definition $(Y_i - \overline{Y})^2$, the solution for r^2 is

$$r^2 = [\Sigma(Y_i - \overline{Y})^2 - \Sigma(Y_i - \hat{Y}_i)^2]/\Sigma(Y_i - \overline{Y})^2 \qquad (3\text{-}32)$$

Finally the numerator can be re-expressed by noting equation (3-27). Thus

$$r^2 = \Sigma(\hat{Y}_i - \overline{Y})^2/\Sigma(Y_i - \overline{Y})^2 \qquad (3\text{-}33)$$

which is the explained variation of Y divided by the total variation of Y .

Complications arise as soon as more than two variables are introduced into the equation. To illustrate, consider a simple three variable example. Thus, of our estimated regression equation is $\hat{Y} = a + bx + cz$, then

$$R^2 = \Sigma(Y_i - \overline{Y})^2/\Sigma(\hat{Y}_i - \overline{Y})^2 \qquad (3\text{-}34)$$

which is the explained variation of Y divided by the total variation of Y . Note that this calculation is identical to r^2 if there is only one independent variable. If there is more than one independent variable, then the numerator represents the variation of Y explained by all independent variables. Thus, as additional explanatory variables are added to the model, we can immediately see how helpful these variables are in improving our explanation of Y by watching how fast R^2 increases in equation (3-34). Finally, it has been proved that equation (3-28) can be generalized in the multiple regression case to:

total variation = variation explained by $(x_1, x_2$ + additional variation explained by x_n + unexplained variation. (3-35)

This statement can be used to construct the ratio "F" = additional variance explained by x_n divided by unexplained variance. (3-36)

It is now appropriate to summarize the differences between the regression and correlation models. The two models differ in the assumptions made about the independent variables. The regression model makes few assumptions about the independent variables, but the more restrictive correlation model requires that the independent variables be random variables, forming with Y a multivariate normal distribution. The regression model may be used to describe the fertilizer-yield problem where fertilizer application is assumed fixed on the one hand, or gives rise to a bivariate normal population of fertilizer and yield on the other. However, the correlation model describes only the latter. It is true that r^2 can be calculated even when fertilizer is fixed, as an indication of how effectively regression reduces variation; but r cannot be used for inferences about the population parameter, ρ. In addition, regression answers more interesting questions. Like correlation, it not only indicates if two variables move together; but also estimates how. Moreover, it can be shown that a key issue in correlation analysis --- the test of the null hypothesis $H_o:\rho = 0$ --- can be answered directly from regression analysis by testing the equivalent null hypothesis $H_o:\beta = 0$. Thus, rejection of $\beta = 0$ implies rejection of $\rho = 0$, and the conclusion must be that correlation does exist between fertilizer and yield. Since regression answers a broader and more interesting set of questions, as well as some correlation questions, it becomes the more comprehensive technique.

To sum up, while simple correlation analysis corresponds to simple regression analysis, the partial correlation analysis corresponds to multiple regression analysis. Recalling how the multiple regression coefficient b estimates how Y is related to x if z were constant, the partial correlation coefficient $r_{xy,z}$ is a similar concept. It estimates the degree to which x and Y move together if z were held constant. Rejection of the hypothesis that $\beta = 0$ is equivalent to rejecting the null hypothesis that $\rho_{xy,z} = 0$. Hence, multiple regression will not only answer its own set of questions, but also partial correlation questions as well.

CHAPTER IV

DATA

Conditions of Sampling

In April 1969, B.C. Tree Fruits Ltd. supplied the Economics Branch, C.D.A., Vancouver and the Department of Agricultural Economics, U.B.C. with current survey data, listing apple tree numbers according to year of planting, rootstock category and variety for individual growers in the Okanagan and Creston Areas of British Columbia.

Three major difficulties in using the survey data for sampling purposes can be cited:

1) Rootstock categories while generally indicating tree size would no doubt fail to do so in the case of intermediate stocks and spur strains (unless growers themselves corrected for this factor).
2) No data were shown for a standard rootstock category. Semi-standard, semi-dwarf and dwarf rootstock categories were included.
3) In showing data for an individual grower no distinction was made between trees in an homogeneous orchard area and trees in an interplanted orchard area.

In order to meet the objectives of this study, it was necessary to select a representative sub-sample of apple enterprises for each of the tree-size categories: standard, semi-standard, semi-dwarf, dwarf.[1] Moreover, the technicalities of costing enterprises made it essential that homogeneous enterprise plots should be selected and costed apart from the rest of orchard fruit on cooperating farms. Thus, from the foregoing explanation, it is obvious that the survey data precluded the ideal population enumeration of growers and enterprises, thereby, considerably restricting sampling sophistication. Nevertheless, it was accepted that in view of there being no alternative data source, the existing survey data could provide an enumeration which although not ideal, would at least lead to a better sample (with time and staff available) than any alternative procedure which dispensed with population data and attempted random selection. This fact will be more appreciated when it is recalled that in 1966, 4,271 census farms were recorded in the Okanagan census division. Most of these were producing apples but only a small proportion were in a position to help with the study.

Sampling Method

The survey data for individual apple growers permitted the following sampling method when broken down by

[1]It should be made clear that tree-size categories reflect the effects of intermediate stocks and spur strains of Scion varieties where these are present. In the common case of just rootstock and scion occurring, tree-size category becomes synonymous with rootstock category.

rootstock categories:

1) Growers were listed according to:

   a) Their having a minimum number (66) or more apple trees in the semi-dwarf category.

   b) Their having a minimum number (100) or more apple trees in the dwarf category, where they had not previously qualified under a) above.

   c) Their having a minimum number (33) or more apple trees in the semi-standard category, where they had not previously qualified under a) or b) above.

Growers who entered these lists were known to be in possession of a minimum number of apple trees of distinct type (corresponding reasonably well with tree size). This would mark them as that much more likely to qualify for sample selection, bearing in mind the high frequency of interplanting and the need to cost individual enterprises of a homogeneous nature with regard to tree-size category, age, density, variety and growing practice. Also it was assumed that growers listed in the manner already explained would make it possible for a sub-sample of standard tree-size enterprises to be selected along with other sub-samples.

2) Within each of the three group lists outlined in 1) above, geographical sub-groupings were made at two levels. Firstly according to N. Okanagan (Westbank and northward), S. Okanagan (southward

from Westbank) and Creston areas, and secondly with regard to constituent districts.

3) District horticulturalists were consulted to make sure that lists of growers referred to managerial entities (i.e., no double counting of a single business structure was permitted). Furthermore, they helped up-date lists whenever it was known that a very recent change in ownership or tenancy had occurred.

4) On the basis of field-worker availability and the inevitable drop-out rate for cooperators, it was decided to obtain 140 apple enterprises for costing in 1969, each one conforming to homogeneity conditions.

Knowledge of apple production in the Okanagan and Creston areas led to the conclusion that 10 enterprises in the dwarf tree-size category would be adequate to represent the small total number of such enterprises. The remaining 130 enterprises were considered best allocated in approximately equal numbers to standard, semi-standard, and semi-dwarf categories.

Since the study required detailed enterprises costings which made it necessary for associated total farm data to be collected, it was decided to limit the total farm accounts to around 100 in order to ensure adequate field-worker time. Up to two enterprises were permitted for each cooperator, although it was correctly deduced that many cooperators would settle for one enterprise.[2]

5) It is now relevant to discuss the purpose of stratification on the basis of enterprise, area and district as referred to in Section 1 - 3 above. The breakdown of apple grower numbers along the lines already described is given below.

Rootstock Category	Grower Population by Area (a)		
	N. Okanagan	S. Okanagan	Creston (b)
Semi-dwarf	109	181	
Dwarf	11+	9+	
Semi-standard	139+	81+	

(a) For Dwarf and Semi-standard categories, the numbers of growers were in excess of figures shown and this is indicated by plus signs. See Section 1. above.

(b) No breakdown for the Creston area is given, since a small number of enterprises, involve 4 growers, was selected by judgment.

It seemed reasonable to expect that sampling within the semi-dwarf and dwarf categories by means of random listing of growers (involving substitution procedure and, if necessary, exhaustion of lists) would achieve random selection of sub-samples across areas for the four categories of apple

[2]The discussion concerns initial selection of enterprises. Later in the study, a few initial single enterprises underwent partitioning to safeguard homogeneity conditions and facilitate analysis. Modification of this type could lead to a grower eventually contributing more than two enterprises.

enterprises.[3] Obtaining more than one enterprise from a grower would not affect randomness providing all growers contacted were given an equal chance of cooperating. Admittedly, the parent population of growers was somewhat reduced in this case and it might be argued more comprehensively in terms of the enterprise constituency shown by individual growers. However, it was still thought satisfactory from the standpoint of useful, statistical inference and it held hopes of being highly efficient in terms of field-work.

Unfortunately, it soon became clear that growers contact within the semi-dwarf and dwarf categories showed a high incidence of either a) inability to help, or b) unwillingness to help, even when possible. In fact, the former was the more important owing to the lack of homogeneous enterprise units. With this experience in mind, the decision was made to extend contacts to the semi-standard category listing. In fact, the very high rate of substitution on randomized listings for the Okanagan areas meant that practically all growers listed were contacted to achieve sufficient numbers of cooperators for each of the sub-samples. Because of the relatively small acreage of apples in the Creston area, a judgment sample of enterprises involving four growers was obtained there with the help of the

[3] If experience showed that exhaustion of lists would be unnecessary, a procedure was devised for maintaining area representation in sub-samples.

District Horticulturist.

The final breakdown of apple enterprises composing the initial total sample drawn in the spring and summer of 1969 was as follows:

TABLE V

Tree-Size Classification of Initial Total Sample

Enterprise Category	No. of Enterprises for Okanagan and Creston Areas
Standard	37
Semi-standard	60
Semi-dwarf	38
Dwarf	7
TOTAL	142

However, it should be made clear that the final sample of apple plots used in the study (n=119) necessitated deletion from the above list where data proved unsatisfactory as well as some partitioning of enterprise data to ensure that homogeneity conditions were met.

An attempt was made to categorize trees into four tree-size groups in accordance with a method suggested by Dr. D. Fisher, Summerland Research Station.[4] This classifi-

[4]It was suggested that four influencing factors, e.g., rootstock, intermediate stock, scion variety and soil type be considered in order that tree size could be represented by an index value. For example, golden delicious on standard intermediate stock on seedling rootstock on poor soil - 1.0x1.0x1.0x0.60 = 0.60. This index value would categorize the above example as semi-dwarf in tree size. Since no accurate information on soil types was available for the study, only the first three factors mentioned above have been taken into account. Further details of deriving the index are shown in Table IV in the Appendix.

cation of apple tree was successfully carried out. However, variety classification was not satisfactorily achieved because a rigorous attempt to group trees into appropriate variety led to arbitrary classification. This frustrating experience stems largely from the fact that a single plot, for example, based on tree-size classification underwent a further partitioning in order to ensure a rigorous variety categorization. Consequently, classification according to variety resulted in the sample size being expanded more rapidly than when classification of tree size was done.

The final sample breakdown of apple enterprises, based on tree-size is shown below:

TABLE VI

Tree-Size Classification of Sample Enterprises

Tree-Size Category	No. of Enterprises for Okanagan and Creston Areas
Standard	23
Semi-standard	62
Semi-dwarf	28
Dwarf	6
TOTAL	119

It should be noted that tree-size categories reflect the effects of intermediate stocks and spur strains of scion varieties where there are present. In the common case of just rootstock and scion occurring, tree-size category becomes synonymous with rootstock category as stated previously.

For each enterprise selected in the study, the following information was obtained:[5]

1. Weight of apple yield
2. Density of apple trees
3. Age of trees
4. Cost of spray applied
5. Cost of fertilizer applied
6. Labour hours spent on pruning and thinning
7. Tree-size index

Since both hired and family labour were employed in pruning and thinning operations, and not all apple producers in the study managed to keep an up-to-date record of labour hours, there is likely to have been some memory bias in recording pruning and thinning hours. In order to calculate data on a per acre basis relevant total enterprise data were divided by corresponding total acreages. Tree age and values of dummy variables representing area differences required no such modification. Implicit in this procedure is an assumption that all independent variables had nothing to do with variation in acreage. The output variable and non-land input variables enter into the analysis as co-efficients or quantities per acre. Independent variables used in the regression analysis are as follows:[6]

1. Apple yield (Y) per acre measured in pounds.
2. Density (D) measured in terms of number of trees per acre (range in study 48 - 605).

---

[5]A copy of the information sheet is presented in the Appendix.

[6]A full list of independent variables (except the dummy variable) is given in Table VII in the Appendix.

3. Age (A) of trees measured in years (range in study 4 - 55). It is of interest to note that apple trees of one to three years of age, which were included in the initial sampling, did not bear any recognizable amount of fruit for the year in which the study was conducted.
4. Fertilizer (F) measured in $ cost per acre. Data regarding the amount of fertilizer used was thought to be less reliable than the cost estimates obtained from growers.
5. Spray (S) measured in $ cost per acre for the same reasons as above. In fact, amounts of spray actually reported were so heterogeneous that it was virtually impossible to derive a meaningful interpretation.
6. Pruning and thinning hours (P) measured in total hours per acre spent on these practices and include hired and other family labour hours.
7. Tree-size index (T) calculated according to a method suggested by Fisher, as explained on Page 38.
8. Dummy variable (G) used for several purposes. The Okanagan area was divided into North and South regions just north of Summerland. This division was made because of environmental differences in the two regions, which were assumed to account for some variation in apple yields. Differences observed between the North and South Okanagan regions include variations in soil type and weather observations for the year

under study. According to the "Climate of British Columbia Report" for 1968 - 1969, slightly different mean temperatures for the two regions were registered during the period May 1968 to May 1969. The average temperatures were 44°F. and 48°F. in the North and South regions respectively. These temperatures were recorded in the growing period, which is defined as the number of days with an average daily temperature above 43°F.[7] The Report also showed that in 1969 there were slight differences between the two regions in precipitation for the months May to October inclusive.

Longley used the May to October period and found there existed a negative relationship between rainfall and apple production for the Annapolis Valley in Nova Scotia.[8] The average precipitation in the northern region of the Okanagan ran from 1.09 to 1.23 inches for the indicated period, whereas in the southern region a relatively low average rainfall of 0.10 to 0.92 inches was reported for the same period.

Variations in yield due to weather factors may be further classified according to direct or indirect action of the causal agent. It is very likely that such weather components as humidity, light and air movements directly

---

[7] The Climate of British Columbia - Tables of Temperature, Precipitations, and Sunshine Report for 1969 - 1970, Province of British Columbia Department of Agriculture. pp. 9-10.

[8] Longley., op. cit., pp. 22-23.

influence yields. Moreover, owing to the relationships and interrelationships present in weather components, yields will be indirectly affected. The intensity of certain insect infestations and plant diseases, for instance, is affected by weather. The effect of weather on apple yield can also vary with the level of fertilizer, soil type, cultural practices, and many other factors. Because of this complexity, the following assumptions were made: (1) non-weather influence is uncorrelated with weather influence; (2) all variations in yield due to non-weather influences are normally distributed with an expected value of zero and a finite variance. Data were split into two parts of approximately equal size. Fifty-four enterprise plots out of a total of one hundred and nineteen were assigned to the north Okanagan region and the remaining sixty-five enterprise plots to the south Okanagan region.

Regional differences in apple yields, as explained earlier, can theoretically be partly explained by dummy variables. A dummy variable is only an indicator variable. It has only two numerical values. In the case of the Okanagan regions '1' was assigned to any enterprise plot in the south and '0' was assigned to any enterprise plot in the north. Modification was necessary in the use of '0' and '1' when the Cobb-Douglas function was used to estimate the yield relationship. The value zero becomes a problem in the process of logarithmic transformations, because Ln 0 approaches $-\infty$. The alternative pair of values, 0.1 and 10, were therefore substituted for '0' and '1', respectively. These two indixes were employed to represent a yield variation, if any, which may be due mainly to differences in locations. Of course, any pair of numbers would serve the purpose equally as well as 0 and 1. But the magnitude of coefficients would vary depending on the values taken by the dummy variables. Hence, interpretation of coefficients derived from a certain pair of numbers is bound to differ from some other pair of numbers.

CHAPTER V

EMPIRICAL RESULTS

Introduction

Before attempting the multiple regression analysis, a simple regression analysis was performed of apple yield on each independent variable, namely, density per acre, age of trees, the cost of fertilizer used per acre, the cost of spray used per acre, and pruning and thinning labour hours per acre. The thing to note is that it seems conceptually very likely that density per acre may be highly correlated with tree-size index. Whether these two independent variables are correlated can readily be checked by the inspection of the correlation matrix. The correlation matrix is given in Table VIII in the Appendix. Notwithstanding the probability of such a correlation occurring, the simple regression of density per acre on tree-size index was tried. Prior to running simple linear regression analyses, data were grouped according to tree-size classification: that is, standard, semi-standard, and semi-dwarf. On these classified data, corresponding simple linear regression analyses were performed with respect to each individual independent variable. Finally, data were lumped together, which permitted simple linear regression analyses to be performed on all overall set of data. In the light of significant regression coefficients and the best 'fit' criterion, the simple linear regression model, regardless of whether data were disaggregated or not, failed to indicate any strong apple yield relationships. The empirical results from the simple linear regression analysis is shown in Table III in the Appendix. The implication from these results may be that apple yield relationships exist with several variables considered simultaneously, and therefore any apple yield relationship might well take a curvelinear form rather than a straight line. This rationale paved the way for multiple regression analysis which is discussed later in the chapter.

The multiple regression routine of the "UBC TRIP" computer program was used to provide least squares regression estimates.[1]

Another program was used for the Equality of Slope Test to see whether the differences in regression coefficients among tree-size groups could be ascribed to sampling errors or to differences among groups.[2] To illustrate, a single variable-of-classification in the form (X_{i1}, Y_{i1}), $(X_{i2}\ Y_{i2})$, (X_{i3}, Y_{i3}) is presented, where X and Y represent fertilizer applied and apple yield, for instance. The first subscript i denotes the number of observations in each group and the

---

[1]J. H. Bjerring and P. Seagraves, UBC TRIP (Triangular Regression Package) Vancouver: U.B.C., Computing Centre, Nov. 1970.

Bill Coshow, UBC BMDX 64: General Linear Hypothesis, U.B.C., Computing Centre, August 1971.

[2]Chinh Le-Dinh, UBC SLTEST: Equality of Slope Test, U.B.C., Computing Centre, June 1971.

second subscript 1, 2, and 3 denote corresponding groups: 1 represents a standard apple group, etc. Naturally these procedures extend to more than a single variable-of-classification. Suppose a question arises as to whether the regression lines corresponding to each group are to be regarded as the same. To answer the question adequately requires construction of the covariance table, as shown below. It will be convenient to denote the quantities in Table IX by individual letters.

TABLE IX

Covariance Table for the Three Tree-Size Groups

Within each group	Σx^2	Σxy	Σy^2	$\Sigma y'^2$
1	C_{xx_1}	C_{xy_1}	C_{yy_1}	C'_{yy_1}
2	C_{xx_2}	C_{xy_2}	C_{yy_2}	C'_{yy_2}
3	C_{xx_3}	C_{xy_3}	C_{yy_3}	C'_{yy_3}
Among means	C_{xxm}	C_{xym}	C_{yym}	C'_{yym}
Within groups	C_{xxw}	C_{xyw}	C_{yyw}	C'_{yyw}
Total	C_{xxt}	C_{xyt}	C_{yyt}	C'_{yyt}

The definitions of the quantities to be computed are as follows:

C_{xx_1}, C_{xx_2}, C_{xx_3} represent the computation $\Sigma X^2 - (\Sigma X)^2/n$ for groups 1, 2, 3.

C_{xy_1}, C_{xy_2}, C_{xy_3} represent the computation $\Sigma XY - \Sigma X \Sigma Y/n$ for groups 1, 2, 3.

C_{yy_1}, C_{yy_2}, C_{yy_3} represent the computation $\Sigma Y^2 - (\Sigma Y)^2/n$ for groups 1, 2, 3.

The quantities in the column $\Sigma y'^2$ are computed by the formula $\Sigma Y^2 - (\Sigma XY)^2/\Sigma X^2$.

The quantities S_1, S_2, S_3, S_4 are defined in terms of C_i' in Table X.

S_1 = the sum of squares of Y values from the regression line in each group, totalled for all groups.

S_2 = the variation among regression coefficients of the different groups

S_3 = the sum of squares of deviations of the means from the regression line of the means with regard to Y values.

S_4 = the square of the difference between coefficients within groups (b_w) and coefficients among means (b_m), and $S_t = S_1 + S_2 + S_3 + S_4$ (see Table XI in Appendix).

TABLE X

Table for S_i in terms of C_i'

Definitions of S_i	D.F.
$S_1 = C'_{yyi}$	$k(n - 2)$
$S_2 = C'_{yyw} - S$	$k - 1$
$S_3 = C'_{yym}$	$k - 2$
$S_4 = C'_{yyt} - C'_{yyw} - C'_{yym}$	1
Total $S_t = C'_{yyt}$	$kn - 2$

Referring to Table X, n = the number of observations and k = the number of groups. Therefore, a test of whether one regression line can be used for all observations can be formulated as follows:

$$F = \frac{\dfrac{S_2 + S_3 + S_4}{2(k - 1)}}{\dfrac{S_1}{k(n - 2)}}$$

The Equality of Slope Test was used for three different tree-size groups, each group containing twenty-seven independent variables. It was used only for the quadratic function because of the priority given to that function as explained in Chapter I.

The results support the hypothesis that there are no differences in corresponding regression coefficients among the groups --- Standard versus Semi-standard, Standard versus Semi-dwarf, and Semi-standard versus Semi-Dwarf. The calculated F = 0.29 and tabulated $F_{.05}$ (D.F.: 30, 55) = 1.67. Approximate values are taken because the Table for the F-test in Snedecor's Statistical Method does not give a value with 35 and 54 degrees of freedom, the ones relevant to the analysis. The results from the F-test are presented in Table XI in the Appendix.

Using tree-size index, the following categorization seemed to be reasonable: Standard tree falling in the range 0.97 - 1.00; Semi-standard in 0.61 - 0.88; Semi-dwarf tree in 0.25 - 0.60; and Dwarf tree in 0.20 - 0.21.

An important thing to note is that no account was taken separately of the dwarf tree group in the study. This group consisted of only six enterprises. As such, it seemed practical to merge the group in with the semi-dwarf group. Further details regarding the Equality of Slope Test will dealt with later. The results of the Equality of Slope tests for tree-size groups were obtained as follows: (1) regression equation for each sample; (2) the twenty-seven common slope coefficients; (3) F-ratio and its probability. This information is given in Table XI in the Appendix. On the evidence of no difference among regression coefficients the three tree-size groups were combined so that a single regression equation might be fitted. Thus, two basic regression models were applied to the overall enterprise data.

The two basic models used in the ensuing regression analysis are as follows: one is a Cobb-Douglas function linear in logarithms; $\text{Ln } Y = \text{Ln } \alpha + \beta_2 \text{ Ln } D + \beta_3 \text{ Ln } A + \beta_4 \text{ Ln } F + \beta_5 \text{ Ln } S + \beta_6 \text{ Ln } P + \beta_7 \text{ Ln } G + \beta_8 \text{ Ln } T + \text{Ln } V$. All that is necessary now is a simple renaming of the terms in this equation:

Y = Ln Y = Yield

β_1 = Ln α

X_1 = Ln D = Log density

X_2 = Ln A = Log age

X_3 = Ln F = Log cost of fertilizer

X_4 = Ln S = Log cost of spray

X_5 = Ln P = Log hours in pruning and thinning

X_6 = Ln G = Log geographical dummy

X_7 = Ln T = Log tree-size index

e = Ln V , where X_1, X_2,----,X_7 represent independent variables used in the study. Assume that V is distributed so that e = Ln V satisfies the assumptions made earlier; namely, Ln V $\sim N(0,\sigma^2)$. Subsequently, the logarithmic equation appears as a familiar linear model. In matrix notation, $Y = X\beta + e$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \cdot \\ \cdot \\ \cdot \\ \\ \cdot \\ Y_{119} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,2}, \text{----}, & x_{1,8} \\ 1 & x_{2,2}, \text{----}, & x_{2,8} \\ 1 & x_{3,2}, \text{----}, & x_{3,8} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \\ \cdot & \cdot & \cdot \\ 1 & x_{119,2}, \text{---}, & x_{119,8} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \cdot \\ \cdot \\ \cdot \\ \beta_8 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ \cdot \\ \cdot \\ \\ \cdot \\ e_{119} \end{bmatrix}$$

Assume that E $(e) = 0$, Cov $(e) = \sigma^2 I$, i.e.

$$\begin{bmatrix} \sigma^2 & & & & & \\ & \sigma^2 & & & 0 & \\ & & \sigma^2 & & & \\ & & & \sigma^2 & & \\ & 0 & & & \sigma^2 & \\ & & & & & \sigma^2 \end{bmatrix}$$

The important requirement of this logarithm transformation is that the error term in the natural form equation is multiplicative. If this assumption is unwarranted, then the model will require special treatment which is beyond the concern of this study.

The other model used is a quadratic function, $Y = \beta_1 + \beta_2 D + \beta_3 A + \beta_4 F + \beta_5 S + \beta_6 P + \beta_7 G + \beta_8 T + \beta_9 D^2 + \beta_{10} A^2 + \beta_{11} F^2 + \beta_{12} S^2 + \beta_{13} P^2 + \beta_{14} DA + \beta_{15} DF + \beta_{16} DS + \beta_{17} DP + \beta_{18} DT + \beta_{19} AF + \beta_{20} AS + \beta_{21} AP + \beta_{22} AT + \beta_{23} FS + \beta_{24} FP + \beta_{25} FT + \beta_{26} SP + \beta_{27} ST + \beta_{28} PT + e$, (where D, A, F, S, P, G, and T refer to density, age, cost of fertilizer, cost of sprays, pruning and thinning hours, geographical dummy and tree-size index respectively). Again, all observations can be stacked into a column vector as follows:

$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $E(\mathbf{e}) = \mathbf{0}$, $\mathrm{Cov}(\mathbf{e}) = \sigma^2\mathbf{I}$.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_{119} \end{bmatrix} = \begin{bmatrix} 1 & x_{1,2}, ----, x_{1,28} \\ 1 & x_{2,2}, ----, x_{2,28} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_{119,2}, ----, x_{119,28} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \\ \beta_{28} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_{119} \end{bmatrix}$$

When one variable is used to obtain several regressors, as in this model, a question may arise as to whether multicollinearity becomes a problem. For example, D_i and D_i^2 are functionally dependent (i.e., one is the square of the other); they are not linearly dependent (i.e., one is not, say, twice the other).

Geometrically, the co-ordinate points (D, D^2) lie on a curve as shown in Figure 2 below; the important thing however, is that they do not lie on a straight line. Thus, the problem of multicollinearity may or may not be avoided according to the degree of curvature involved.

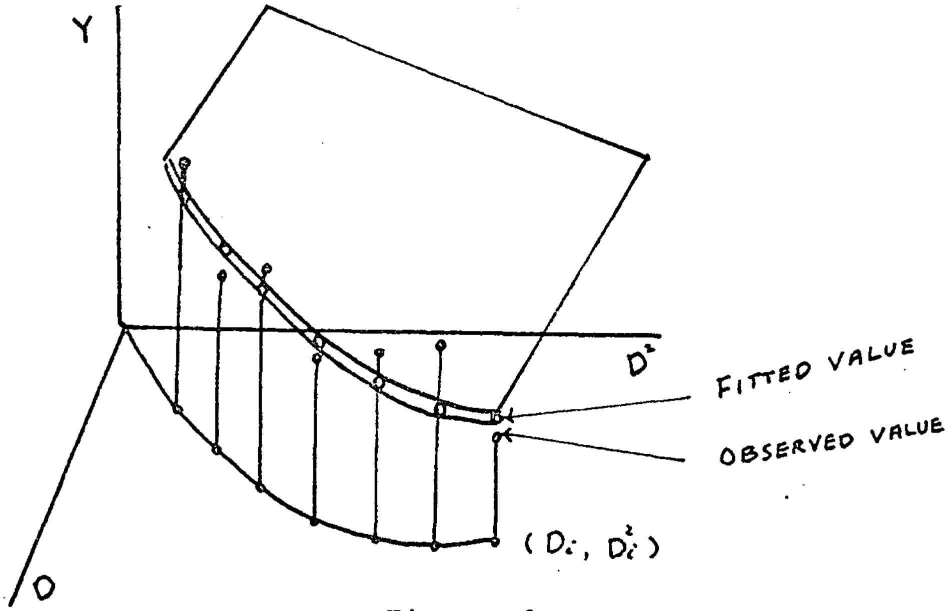

Figure 2.
Polynomial regression as a special case of multiple regression.

The output of the Trip program for both regression models included: (1) the estimated regression coefficients; (2) the standard error of each coefficient; (3) the F-ratio and associated probability for each regression coefficient; (4) the standard error of the estimate, $\hat{Y}$; (f) the coefficient of multiple determination, R^2; and (6) the correlation matrix.

In showing data, subsequently standard errors of the regression coefficients are shown in parenthesis. The associated probability of the F-ratio for each coefficient is shown below each standard error. The results of the estimated stepwise Cobb-Douglas regression equation are given in Table XII in the Appendix and the correlation matrix for the Cobb-Douglas function is shown in Table XIII in the Appendix. The results of the estimated quadratic model are presented on page 57 and the step-wise regression equation is shown in Table XIV in the Appendix. Its correlation matrix appears in Table XV in the Appendix.

Results from the Cobb-Douglas Model

The estimated overall enterprise regression equation in logarithms is:

$$\hat{Y} = \underset{(1.4674)}{1.5514} + \underset{(0.2331)}{0.5713}\,D + \underset{(0.2226)}{1.2263}\,A + \underset{(0.0965)}{0.22263}\,F + \underset{(0.1053)}{0.2485}\,S +$$

$$\underset{(0.0748)}{0.1967}\,P + \underset{(0.0423)}{0.1069}\,G + \underset{(0.2455)}{0.1712}\,T\,. \quad R^2 = 0.4147, \text{ where}$$

all variables are expressed in logarithmic form.

The regression coefficients for all variables except that for tree size were found significantly different from zero at the 5% level of probability. Approximately 40% of total variation in crop yield (Y) has been accounted for by the independent variables.

The value of R^2 is not improved in the stepwise regression equation when only those independent variables which make a significant contribution to apple yield are included. The results of the stepwise regression and corresponding correlation matrix are shown in Table XII and XIII respectively in the Appendix.

In view of the fact that primary interest in the apple study is in the regression model rather than the correlation model, the multiple correlation coefficient R cannot be considered strictly as an estimate of the population correlation between the independent variables and the dependent variable. This is because the independent variables in the regression model are observed in terms of given values and not a multivariate normal distribution. Even so, R does provide a summary statistic to measure the goodness of fit of the observed points to the regression plane.[3]

Results from the Quadratic Model

Before proceeding with the combined data, several points should be made in order to clarify the underlying concepts involved in the employment of Equality of Slope Test.

Firstly, the Test is part of "Analysis of Covariance", the primary concern of which is to find out whether a single regression line is statistically valid in representing a yield relationship. Consequently, the analysis does not produce R^2 values. Secondly, the analysis is incapable of automatically eliminating the insignificant variables and performing the Equality of Slope Test with only the remaining significant variables. Thirdly, the most difficult problem is in deciding which variables are to be retained, and which are to be omitted from the model.

Generation of innumerable terms from variables

---

[3]M. Ezekiel and K. A. Fox, Methods of Correlation and Regression Analysis, 3rd edition. pp. 270-281, 1963.

squared or combinations of the seven basic independent variables is possible: G^2, T^2, GT, etc. But from the subjective point of view, the variables which have been excluded would seem to be lacking any logical basis, variables included are justified in that they are capable of helping to represent a biological phenomenon, i.e., the law of diminishing returns in the case of the squared terms. Even so, squared terms like G^2 and T^2 can in no way appeal to the senses by which subjective judgment is made. For the same reason, some of the cross-terms do not appear in the model.

The "UBC SLTEST" was used for the purpose of Equality of Slope Test. The result of the test with respect to each regression equation is as follows:

1. Regression Equation for Standard Tree-Size Group.

$$\begin{aligned}\hat{Y} = {} & 989.300 + 1212\ D + 15.060\ A + 1457\ F + 3374\ S + 2742\ P \\ & + 1466\ G - 1043\ T - 29.360\ D^2 - 235.600\ A^2 + 18.130\ F^2 \\ & + 2.353\ S^2 - 2.143\ P^2 - 118.700\ DA + 55.200\ DF + 15.330\ DS \\ & - 24.650\ DP + 7484\ DT + 95.960\ AF + 62.690\ AS - 30.760\ AP \\ & - 12.980\ AT - 7.246\ FS - 2.846\ FP - 8113\ FT + 2.474\ SP \\ & - 7018\ ST + 11.130.\end{aligned}$$

2. Regression Equation for Semi-Standard Tree-Size Group.

$$\begin{aligned}\hat{Y} = {} & - 7.650 - 668.4000\ D + 2.9720\ A - 4246\ F + 2872\ S \\ & - 2515\ P - 878.4000\ G - 2.2360\ T - 0.8332\ D^2 - 265.7000\ A^2 \\ & + 15.6500\ F^2 + 1.3470\ S^2 + 2.5170\ P^2 + 1.7850\ DA - 9.5550\ DF \\ & - 6487\ DS + 9.4690\ DP + 1263\ DT + 24.4000\ AF - 36.0300\ AS \\ & + 17.3800\ AP - 2.7030\ AT + 30.4000\ FS - 15.6700\ FP\end{aligned}$$

$$+ 5115\ FT - 9.0710\ SP - 2186\ ST + 1863\ PT\ .$$

3. Regression Equation for Semi-Dwarf Tree-Size Group.

$$\hat{Y} = - 12.5100 - 162\ D + 2793\ A - 1895\ F = 781.200\ S + 1057\ P$$
$$+ 491.500\ G + 50.2600\ T + 0.6151\ D^2 + 202.7000\ A^2$$
$$- 3.264\ F^2 + 3.6610\ S^2 - 0.1218\ P^2 + 11.7900\ DA + 1.9570\ DF$$
$$- 1.1880\ DS - 0.9559\ DP - 418.6000\ DT - 57.0900\ AF$$
$$+ 38.4700\ AS - 29.0100\ AP - 2.445\ AT - 8.5100\ FS$$
$$- 5.8330\ FP + 8298\ FT + 1.6100\ SP - 3448\ ST - 1442\ PT\ .$$

An F-test was performed on the twenty-seven coefficients held in common by each regression equation. The result indicates that there are no significant differences in comparable regression coefficients among the three tree-size groups at the 5% level of significance.

Data from the test are presented in Table XI in the Appendix. Therefore on the basis of this result the three separate samples were combined into one sample. A quadratic regression equation was then estimated as follows:

$$\hat{Y} = - \underset{(7.818)}{3.200} + \underset{(210.5007)}{96.5140\ D} + \underset{(4881.8270)}{3527.7863\ A} - \underset{(984.6400)}{306.9661\ F} + \underset{(451.9521)}{365.2424\ S}$$
$$- \underset{(379.0393)}{698.2985\ P} - \underset{(665.2497)}{231.1846\ G} + \underset{(8.189)}{3.7550\ T} - \underset{(0.2094)}{0.004814\ D^2} - \underset{(32.5372)}{18.7206\ A^2}$$
$$- \underset{(3.8344)}{6.9504\ F^2} - \underset{(1.1533)}{1.5232\ S^2} + \underset{(0.3758)}{0.6457\ P^2} + \underset{(11.9919)}{6.5565\ DA} + \underset{(2.1995)}{1.4649\ DF}$$
$$- \underset{(1.1785)}{1.3623\ DS} + \underset{(0.8984)}{1.1930\ DP} - \underset{(166.1315)}{218.4573\ DT} - \underset{(54.1606)}{26.8412\ AF} - \underset{(0.9068)}{3.0749\ AS}$$
$$- \underset{(10.5395)}{2.0330\ DS} + \underset{(4345.6588)}{2610.1141\ AT} + \underset{(3.4245)}{6.1613\ FS} + \underset{(3.2792)}{0.1162\ FP} + \underset{(901.3124)}{1004.4861\ F}$$
$$+ \underset{(0.8558)}{0.5085\ SP} + \underset{(414.9117)}{170.2734\ ST} + \underset{(344.7191)}{511.4871\ GT}\ . \quad R^2 = 0.7534$$

The fact that eleven regressor coefficients in the above equation were not significant at the 5 per cent level can immediately be checked by observing that the standard errors in parentheses of the coefficients exceeded values of the corresponding coefficients. Non-significance is also true of other coefficients in the equation but stepwise regression at a later stage will select the significant variables for a final equation analysis.

A question at this stage may arise as to why this kind of situation has occurred. The first necessary step to take is to examine whether any of the appropriate assumptions made in connection with estimating the quadratic function have been violated. Therefore, initially the correlation matrix[4] must be investigated to see if multicollinearity might have caused problems. A close inspection shows that there are a number of near-linear combinations formed between independent variables and regressors (no linear dependence was shown between independent variables) most of which have been generated in the process of either squaring an independent variable or interacting one independent variable with another. These occurrences are a direct violation of the assumption that a regressor D^2, for example, is functionally but not linearly dependent on D. However, if the curve segment on which the coordinate point D_i and D_i^2 in Figure 2 lies is close to the shape of a straight line segment, there can be problems of multi-

[4]See Table XV in the Appendix.

collinearity.

If D_i and D_i^2 form an almost near-linear combination, the variable $x_{1,2}$ and $x_{1,9}$ in terms of matrix will be almost linearly dependent:

$$\mathbf{X} = \boxed{\begin{matrix} 1 & x_{1,2}, ----, & x_{1,9}, ----, & x_{1,28} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{119,2}, ----, & x_{119,9}, ----, & x_{119,28} \end{matrix}}$$

Such multicollinearity results in extremely large entries in the inverse matrix $(\mathbf{X}'\mathbf{X})^{-1}$. Since $\sigma^2\ (\mathbf{X}'\mathbf{X})^{-1}$ is the covariance matrix for the $\hat{\beta}$,[5] we therefore obtain very large covariances, and hence broad confidence intervals.

The multicollinearity problem may be clearly visualized, geometrically, in Figure 3. But to keep the geometry manageable, an ellipsoid that delimits most of the β_1's, the so-called "ellipsoid of concentration" is shown. For the independent errors assumed earlier, the ellipsoid is simply a sphere. This sphere of Y observations is centered at the mean E ($\mathbf{Y}$), which is in the plane generated by $\mathbf{X}_1$ and $\mathbf{X}_2$. Figure 3 shows what happens when regressors $\mathbf{X}_1$ and $\mathbf{X}_2$ are not orthogonal mutually (perpendicular) but collinear, the interval of $\hat{\beta}_1$'s is dispersed on both sides of the origin. The point estimate may be positive, but there is a good chance it may be negative.

[5] J. Johnston, Econometric Methods, New York: McGraw-Hill, p. 110, 1960.

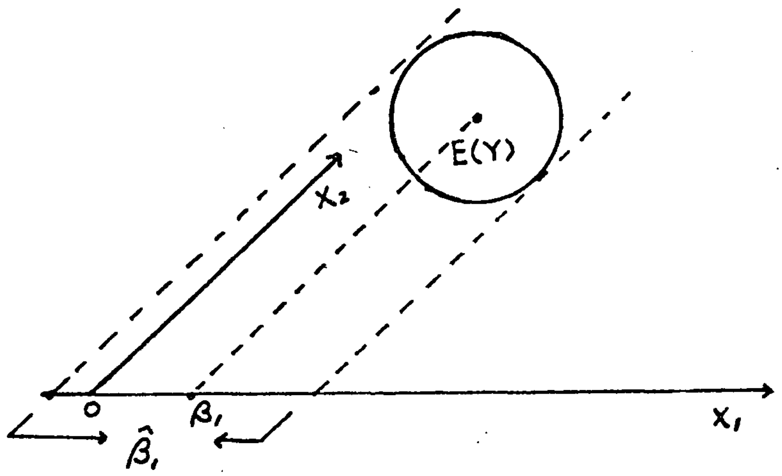

Figure 3
Range of values for possible $\hat{\beta}_1$'s around origin when x_1 and x_2 are highly collinear.

Although Figure 3 shows that the true β_1 is not zero, this is very difficult to establish statistically. Usually $H_0(\beta_1 = 0)$ will not be rejected under conditions where there is a huge standard error of $\hat{\beta}_1$.

If, therefore, any values of these regressors in the correlation matrix are close to, say, $|0.8|$, the regression analysis should be carried out with one of the highly correlated variables omitted. It is, however, extremely difficult to decide which regressors to omit and which to retain because those regressors included in the equation have been selected on the basis of logic --- physical or biological --- relevant to the production process being examined. Under these circumstances it is possible to test, by stepwise regression, whether or not each of the regressors (and for that matter other independent variables) is making a significant contribution to explaining variation in yield. The forward stepwise regression quadratic equation actually selected the following variables at the 5 per cent level of significance:

$$\hat{Y} = 5733.2578 + 2739.7249\ A + 398.6005\ S - 1096.3665\ P - 8.6445\ F^2 + 0.9406\ P^2 + 3.5579\ DF - 2.2731\ DS + 1.9925\ DP - 2671.2339\ AT + 2.8387\ FS + 866.9939\ PT\ .$$

$R^2 = 0.7212$.

Once multicollinearity becomes a problem, even stepwise regression would not help resolve it. Stepwise regression is designed to select independent variables least linearly combined in the first place, and next less linearly combined and so on in the order of independent variables laid out in regression equation. It follows that forward regression does not necessarily coincide with backward regression if independent variables are collinear. Therefore, selected independent variables may differ according to the regression routine insturction, i.e., forwards or backwards.

Moreover, if some independent variables are linearly dependent on the other ones, the value of R^2 becomes dubious. Coordinate points of linearly-dependent independent variables are not spread out but clustered in nearly linear fashion in dimensional space involved, and thus the determination of a meaningful regression surface by least squares method is rendered that much more difficult.

There is a point that should be made about the distribution of the apple-yield dependent variable with respect to fixed values of an independent variable in the regression equation. From the fact that the error term e is assumed normally distributed with mean = 0 and variance $= \sigma^2$, it follows that the random variable, apple yield, for specified values of the independent variables, is also assumed normally distributed with mean $= \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7$ and variance $= \sigma^2$, if a regression model in regard to apple production is constructed in the following manner: $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + e$. One hundred and nineteen observations are large enough to validate this assumption. If, however, the assumption is not met, parametric methods which have been employed are no longer adequate. There are two alternative ways to deal with this situation. One is perhaps to transform the raw data using logarithmic or square-root transformations, etc. The other way to cope with this problem is to apply non-parametric statistics in which case the technique is entirely beyond the scope of the study.

The details of results from stepwise regression are tabulated and shown in Table XIV in the Appendix. Table XVI in the Appendix shows both observed values and the corresponding values of apple-crop yield on the basis of the above equation involving the selected independent variables with significant regression coefficients.

Discussion of the Results from Applying Cobb-Douglas and Quadratic Regression Analyses

On a priori grounds, the results of the Cobb-Douglas function show expected signs for the regression coefficients of the seven regression coefficients estimated. Only that of the tree-size variables was not found significant at the 5 per cent level of probability. Using the quadratic function, only the three basic independent variables, age of tree; cost of spray; and pruning and thinning hours were significant at the 5 per cent level.

While the Cobb-Douglas function in this case does not produce a multicollinearity problem, the quadratic function has shown evidence of multicollinearity as indicated by inspection of the correlation matrix. If the multicollinearity condition exists between two variables, it is very difficult to establish the level of statistical significance of coefficients. Therefore the influence on crop yield of one variable may be erroneously attributed to the other.

It would seem reasonable to say that this study provided insufficient evidence for choosing between the two models. Preference for the quadratic function over the Cobb-Douglas may be stated on the purely deductive or theoretical grounds that interaction of factors can be at work. However, this kind of selection of the quadratic function is made on the same grounds as explained by Hume's philosophical insights:

"When we give the preference to one set of arguments above another, we do nothing but decide from our feeling concerning the superiority of their influence."

CHAPTER VI

TESTING FOR THE DIFFERENCE BETWEEN TWO MEANS

Introduction

In conducting the tests it was necessary to use overall unweighted per-acre means for specified groupings of apple plots. For instance, since the sizes of apple enterprises were initially determined more by convenience of record keeping than by the relationship they have to the total acres of enterprises on the farms (not readily definable), it became practicable to calculate a per-acre average for each variable on each plot (the unweighted mean). These could then allow an overall per-acre mean to be calculated for a particular variable across any sample group. While these types of average are permissible, their interpretation is limited strictly to the narrow context in which they were derived. Hence, the tests may show differences or no differences among means, but it must be remembered that the means relate to individual farmer performance in apple yields where each farmer has a weight of one. Also it is important to realize that a difference is with regard to the samples for 1969 and these samples may show quite different distributions with regard to age, density, cost of fertilizer, cost of spray, pruning and thinning labour hours, tree-size index, and geographical location. Therefore, the difference or lack of difference in the type of means used can easily be seen as a result of influences of the above variables. In conclusion, it can be said that the results bear very careful interpretation and are to be seen only as slightly extending our insights into the yield relationship picture already studied in the previous chapters.

A t-test is used to test the null hypothesis that two samples come from populations with the same mean: consequently, this tests whether two samples are "significantly different" in this regard. The ordinary method of making a test of significance for the difference between means of two independent samples assumes that the two population variances are the same.[1]

It has been assumed about the apple crop yield Y, that a sample mean $\overline{Y}_1$, is normally distributed around the population mean, μ, as follows: $\overline{Y}_1 \sim N(\mu_1, \sigma^2/n_1)$, where σ^2 represents the variance of the population, and n the size of the sample drawn. Similarly, $\overline{Y}_2 \sim N(\mu_2, \sigma^2/n_2)$. Independence of the two sampling procedures will ensure that the two random variables $\overline{Y}_1$ and $\overline{Y}_2$ are independent: $(\overline{Y}_1 - \overline{Y}_2) \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$.

When population variance σ^2 is unknown, it must be estimated:

$$S_p^2 = \left(\frac{1}{n_1 + n_2 - 2} \right) \left[\left(\Sigma_{i=1}^{n_1} (X_{1_i} - \overline{X}_1)^2 + \Sigma_{i=1}^{n_2} (X_{2_i} - \overline{X}_2)^2 \right) \right],$$

where S_p^2 = pooled variance. The formula for the t-test is:

[1]G. W. Snedecor and W. G. Cochran, Statistical Method, 6th Ed., pp. 114-115, 1969.

calculated $t = \bar{Y}_1 - \bar{Y}_2 / \sqrt{S_p(1/n_1 + 1/n_2)}$.

There may exist situations in which the assumption that $\sigma_1^2 = \sigma_2^2$ is suspect. If so, the formula for the variance of $(\bar{Y}_1 - \bar{Y}_2)$ in independent samples still holds, namely, $\sigma^2_{\bar{y}_1 - \bar{y}_2} = \sigma_1^2/n_1 + \sigma_2^2/n_2$. When σ^2 is unknown, the unbiased estimator S^2 is substituted. The ordinary t value is replaced by the statistic: $t' = \bar{Y}_1 - \bar{Y}_2 \big/ \sqrt{S_1/n_1 + S_2/n_2}$. This quantity does not follow student's t-distribution when $n_1 = n_2$.[2] If, however, sample size of each group is equal, t and t' become identical. On the other hand, if the samples are not of equal size, only approximate degrees of freedom will be calculated by the following formula:

$(S_1^2/n_1 + S_2^2/n_2)^2 \big/ [(S_1/n_1)^2/n_1\text{-}1 + (S_2/n_2)/n_2\text{-}2)]$.[3]

It should be noted that the yield measurements are obtained on a per-acre basis, and t-tests throughout are performed on this basis.

As mentioned earlier in the thesis, sampling was not conducted from the three separate tree-size group populations. Rather, sub-samples of tree-size and other group enterprises were obtained from the sample of apple producers drawn from a single population. In the analyses which follow, sample sizes of groups are unequal but this feature is permitted by the computer program.

The t-test routine of the TRIP program has three

---

[2]Ibid. p. 115.

[3]R. E. Walpole, Introduction to Statistics, The MacMillan Co., Collier-MacMillan Limited, London. pp. 230-231, 1968.

different formulae at its disposal:

Formula (1): the only assumption made about the parent populations in the derivation of this formula is normality.

Formula (2): this is a special formula used when there are differences in the data paired scores (of no concern in this study).

Formula (3): this is a more sensitive version of Formula (1). Formula (3) is valid only when the population variances are equal.

In fact, users can request the t-test to use Formula (1) if it finds the sample variances significantly different, and to use Formula (3) when that is not the case.

Outcome of T-test

Before showing a test for average yield differences, it is desirable to have a picture of differences among average yields for the specific categories studied. The respective figures below are presented to serve this purpose and following each has a table showing details of the relevant test.

Tree-Size Group | Average Apple Yield by Tree-Size Group | Sample Size

Standard

Semi-Standard

Semi-Dwarf

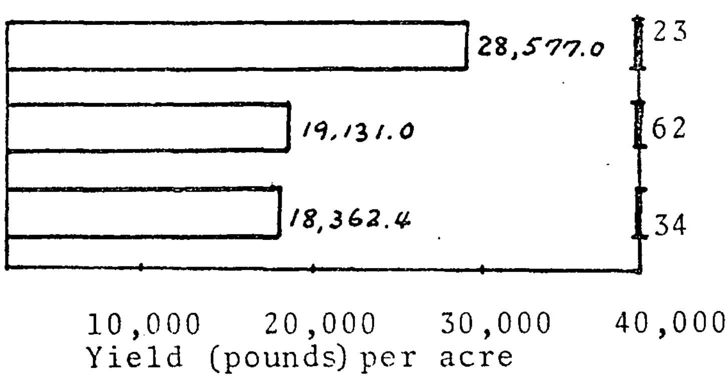

Figure 4

Differences in average apple yields among tree-size groups.

The results from t-tests concerning Figure 4 are shown in Table XVII.

TABLE XVII

Results from t-tests for average-apple-yield differences relating to tree-size groups.

Tree-Size Group	Calculated T-value	D.F.	T-Prob.	F-Prob.	Formula Used
Standard vs. Semi-Standard	2.132	83	0.034	0.163	(3)
Standard vs. Semi-Dwarf	2.450	55	0.016	0.730	(3)
Semi-Standard vs. Semi-Dwarf	0.199	94	0.823	0.225	(3)

If the T-Probability is less than 0.05, it is usually concluded that the sample means are significantly different. If the F-Probability is less than 0.05, it is usually concluded that sample variances are significantly different and therefore formula (3) is inappropriate for calculating t. According to these criteria, the average apple yield difference between Standard and Semi-Standard tree-size groups is significantly different. The same is true between Standard and Semi-Dwarf tree-size groups. But this was not the case with Semi-Standard and Semi-Dwarf tree-size groups where the difference between means was found not to be significant. The three corresponding pairs of sample variances were found not to show significant differences and thus formula (3) was used throughout the t-test.

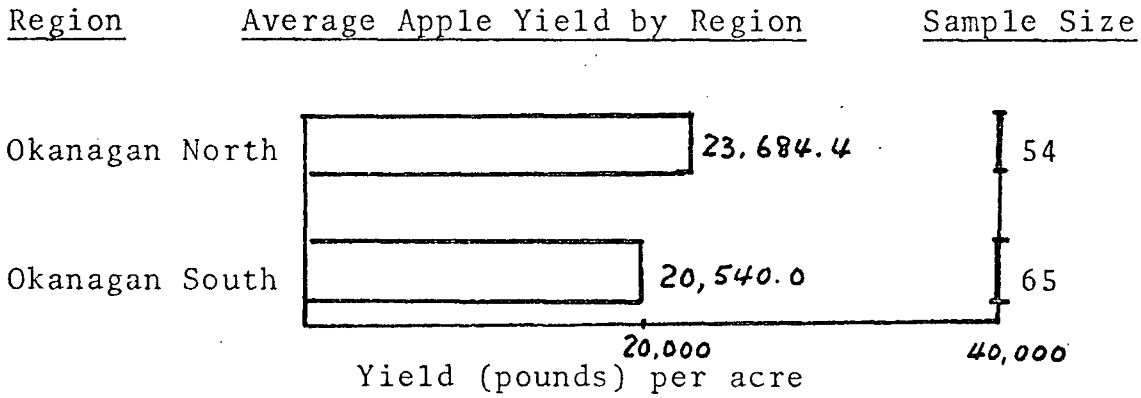

Figure 5

Difference in average apple yields between regions (across all tree-size groups)

The outcome of the t-test concerning Figure 5 are shown in Table XVIII.

TABLE XVIII

Results from t-test for average-apple-yield difference between regions

Regions	Calculated T-value	D.F.	T-Prob.	F-Prob.	Formula Used
Okanagan North vs. Okanagan South	0.581	73	0.570	0.0	(1)

There is no significant difference in average yield between the Okanagan North and Okanagan South regions. But their sample variances are significantly different.

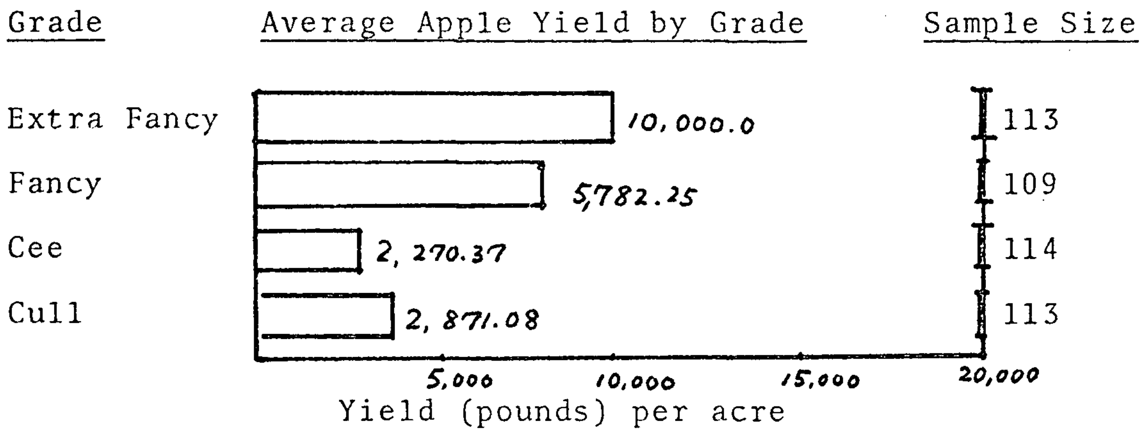

Figure 6

Differences in average apple yields among apple grades (across all tree-size groups)

The outcome of the t-test concerning Figure 6 are shown in Table XIX.

TABLE XIX

Results from t-tests for average-apple-yield differences related to grades

Grade	Calculated T-value	D.F.	T-Prob.	F-Prob.	Formula Used
Extra Fancy vs. Fancy	3.353	159	0.001	0.0	(1)
Extra Fancy vs. Cee	6.472	134	0.0	0.0	(1)
Extra Fancy vs. Cull	5.367	184	0.0	0.0	(1)
Fancy vs. Cee	5.448	191	0.0	0.1	(1)
Fancy vs. Cull	3.354	210	0.001	0.006	(1)
Cee vs. Cull	-0.777	170	0.444	0.0	(1)

The differences between average yields for pairs of grades are significant with the exception of Cee versus Cull. The sample variances are also significantly different for all pairs of grade couplings.

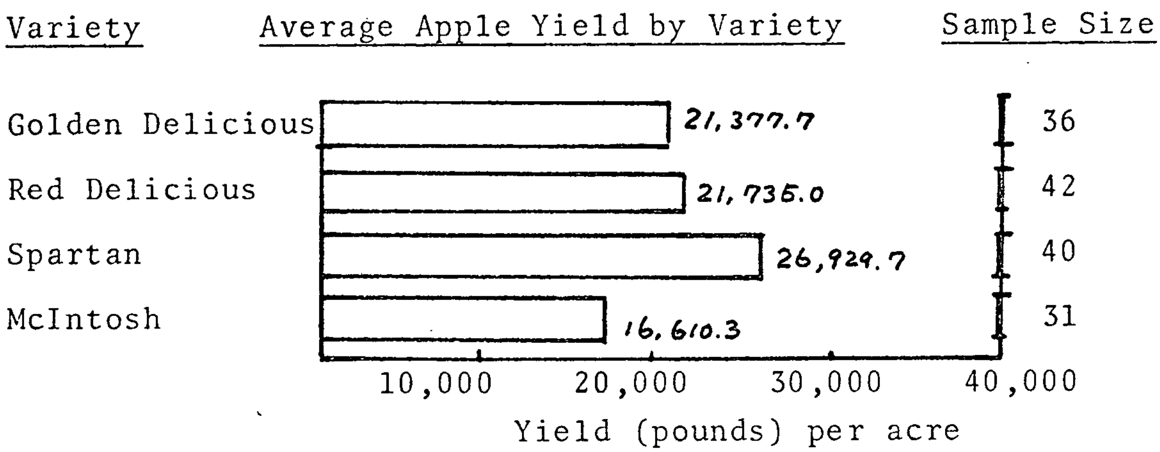

Figure 7

Differences in average apple yields among apple varieties (across all tree-size groups)

As has been mentioned in the introductory chapter, the breakdown of data into varieties has resulted in an arbitrary manipulation of data. When sub-sampling was carried out, it was done in accordance with the tree-size groups, and hence a single enterprise could sometimes involve different kinds of apple varieties.

The rigorous attempt to group data by variety has contributed to enlarged sample size in some instance simply because it involved partitioning original enterprises. In this procedure the underlying assumptions of ensuring analysis were not thought to be infringed seriously, although some caution in acceptance of the results is thought necessary. The results of the t-tests on variety yield data are given in Table XX.

TABLE XX

Results from t-test for average-apple-yield differences relating to variety

Variety	Calculated T-value	D.F.	T-Prob.	F-Prob.	Formula Used
Golden Delicious vs. Red Delicious	-0.084	76	0.893	0.552	(3)
Golden Delicious vs. Spartan	-0.747	56	0.465	0.0	(1)
Golden Delicious vs. McIntosh	1.166	62	0.247	0.045	(1)
Red Delicious vs. Spartan	-0.719	52	0.482	0.0	(1)
Red Delicious vs. McIntosh	1.336	71	0.182	0.129	(3)
Spartan vs. McIntosh	1.451	49	0.149	0.0	(1)

Varietal differences in mean values of crop yield are not significant. Furthermore, the sample variances for Golden Delicious and Red Delicious are not significantly different, as was the case also for Red Delicious and McIntosh. For all other pairs of varieties, the variances were significantly different.

Discussion of the t-test

The significant differences in average yields between the Standard and Semi-Standard groups and between the Standard and Semi-Dwarf groups was surprising because the tree-size index variable had been found not statistically significant as was dropped from the regression equations.

The reason for this may be found in the interpretation of the characteristics of the t-test in relation to regression analysis. The t-test may be seen as a simple regression on dummy variables representing the three different tree-size groups. The following situation can be depicted:

	Dummy variables		
Standard group	0	0	0
Semi-Standard group	1	0	0
Semi-Dwarf group	0	1	0

A dummy variable for the Standard group is not needed because Semi-Standard and Semi-Dwarf groups reflect differentials measured from the Standard group base. This situation can easily be visualized in matrix notation:

$$\begin{bmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \cdot \\ \hat{Y}_{119} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \cdot & \cdot & \cdot \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

Dummy variables are not the only means of adjusting data. Another method is to devise a scale for tree-size as has been done earlier in this study. The desired relation of crop yield to the tree-size variable can now be estimated by a simple regression of apple yield on tree-size index. This latter method is advantageous because it is not necessary to assume discrete shifts. Thus adjustment for any observation varies from one observation to another. This same idea was utilized when hypothesizing the Equality of Slopes among the tree-size groups. It is crucial to remember that the production relationships under examination by the t-test are theoretically implied within the framework of the multiple regression model using seven independent variables and twenty regressors.

The t-test of the type used has limited application in the data context of the study. It is a meaningful method deriving information under the condition of single variable-classification where other influences are held constant. It is precisely our inability to hold other influences constant when testing the yield differences for pairs of categories, which renders the results from the t-test less powerful than those obtained from multiple regression. Nevertheless the tests are of interest because they do provide further insights by way of either affording a slightly different attack or even adding to the overall analysis, e.g., differences among grades.

CHAPTER VII

SUMMARY AND CONCLUSION

The highest achievement would be to grasp that whatever we call a "fact" is already theory. Goethe

Summary

The objective of this study was to estimate regression relationships between apple yields and certain influencing factors for the Okanagan area of British Columbia in 1969.

Two types of equation were employed to represent yield relationships. These were the Cobb-Douglas and Quadratic forms.

The explanatory variables used in regression were as follows: (1) density per acre, (2) age of tree, (3) the cost of fertilizer applied per acre, (4) the cost of spray applied per acre, (5) pruning and thinning labour hours per acre, (6) geographical dummy variable, and (7) tree-size index.

When a Cobb-Douglas function was fitted to all sample observations (across tree-size groups), the independent variables were found significant at 5 per cent level of probability with the exception of tree-size index. On the other hand, a Quadratic function involving twenty-eight independent terms, included only eleven terms as being significant at the 5 per cent level of probability. Here the selected variables were as follows: (1) age of tree, (2) cost of spray per acre, (3) pruning and thinning labour hours per acre, (4) squared fertilizer cost per acre term, (5) squared pruning and thinning hours per acre term, (6) cross-term between density and fertilizer cost per acre, (7) cross-term between density and spray cost per acre, (8) cross-term between density and pruning and thinning labour hours per acre, (9) cross-term between age of tree and tree-size index, (10) cross-term between fertilizer and spray costs per acre, and (11) cross-term between pruning and thinning labour hours per acre and tree-size index.

In view of the properties of a Quadratic function and the economic theory of production, a choice was made in favour of it relative to the Cobb-Douglas function to represent the apple-yield relationship. But the Quadratic function resulted in a complicated problem arising from squaring independent variables to be used as regressors. The use of an independent variable along with its squared term may have generated some collinearity. If two independent variables cause a multicollinearity problem it is extremely difficult to deduce the influence of one of the variables on the dependent variable because it might well be that the other variable has equal influence.

Conclusion

Looking back to the regression model, several assumptions were made so that inferences from estimated regression equations could be made.

Suppose that the following functional relationship exists in regard to apple yield $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + e$. A strong assumption must be made about e: namely, $e \sim N(0,\sigma^2)$. In this model it is implied that all independent variables are treated as fixed. The only random variable in the model is Y which is deduced from the fact that e is a random variable.

Another important assumption made in the model is that all independent variables are independent of one another, and of the error term e. Although the terms "random variable" and "independence" test our powers of comprehension, they do in fact correspond to empirically determinable features in certain actual processes as a consequence of various rules employed by statisticians.

Treatment of all independent variables as taking fixed values implies that the only error allowed was an error in the equation due to the omission of some input factors. In fact, error in the measurement of the included input variables is extremely likely. The error may be due to "human element" involved --- for instance, mistakes may occur in the collection and recording of data.

The observed values of the variables are not strictly comparable because of lack of homogeneity which is the case with fertilizer cost, spray cost, and pruning and thinning labour hours. Therefore observation errors are necessarily present in the data. It is possible that some method of adjusting the data to take account of heterogeneity might be used, but even so it is difficult to contrive and most certainly would leave something to be desired. In the case of the study it is assumed that measurement error is not of serious proportion.

As long as the independent variables are not orthogonal, $X_1 \cdot X_2 = 0$ for example, stepwise regression cannot yield a satisfactory result.[1]

In view of the methodological issues in the foregoing statistical analysis, it is appropriate to say that the empirical results obtained are inconclusive. In this connection it is to be suggested that to meet all the assumptions required for deriving an apple yield equation, a controlled experiment would be necessary. Within the theoretical framework, preference may be given to the Quadratic form of yield relationship over that summarized by a Cobb-Douglas function. However, the former function has posed a serious statistical problem (multicollinearity), which has already been discussed at some length.

Needless to say, it is hoped that any future study under similar circumstances to the one which has been con-

---

[1]Wonnacott, et al., Econometrics, pp. 309-312.

ducted will be in a better position to use the Cobb-Douglas function over selected relevant ranges of data and, thereby, avoid the multicollinearity problem as met in the Quadratic analysis.

Since the regression theory gives rise to the most difficult conceptual part of the thesis, this brief summary and conclusions has helped to remind the reader of the very real obstacles to the type of analysis undertaken. In the case of the tests of significance for differences among means, the conceptual framework is somewhat easier to understand, although the analysis rests on definite assumptions. These as well as the results of that analysis occur in a preceding chapter, and therefore no attempt is made to repeat the summary already given.

BIBLIOGRAPHY

BIBLIOGRAPHY

Bjerring, J. H. and Seagraves, P. UBC TRIP (Triangular Regression Package) Vancouver: U.B.C., Computing Centre, November 1970.

Brase, K. D. and Way, R. D. "Rootstocks and Methods used for Dwarfing Fruit Trees," New York State Agricultural Experiment Station, 783, 1959.

Bush, R. Tree Fruit Growing, revised by E. G. Gilbert prepared in conjunction and collaboration with the Royal Horticultural Society, Penguin Books, 1962.

Chinh Le-Dinh. UBC SLTESTS: Equality of Slope Test, U.B.C., Computing Centre, June 1971.

Coshow, W. UBC BMDX64: General Linear Hypothesis, U.B.C., Computing Centre, August 1971.

Doll, J. P. "An Analytical Technique for Estimating a Weather Index from Meteorological Measurements," Journal of Farm Economics, Vol. 49, No. 1, February 1967.

Dorling, M. J. The Okanagan Apple Producer --- His Management Attitude and Behaviour, Department of Agricultural Economics, U.B.C., 1968.

Ezekiel, M. and Fox, K. A. Methods of Correlation and Regression Analysis, 3rd Ed., 1963.

Fisher, D. V. High-Density Orchards for British Columbia Conditions, Research Station Summerland, British Columbia Research Branch, Canada Department of Agriculture, March 1966.

Folger, J. C. and Thomson, S. M. The Commercial Apple Industry of North America, Ed. L. H. Baily. New York: MacMillan Co., 1921.

Haavelmo, T. "The Probability Approach in Econometrics," Econometrica, Vol. 12, 1944. Supplement.

Harris, J. H. and Woods, J. J. Dwarfing Apple Trees on Vancouver Island, Experimental Farm Research Branch, Saanichton, B. C., 1958.

Heady, E. O. and Dillon, J. Agricultural Production Functions, Ames, Iowa: Iowa State University Press, 1961.

Johnston, J. Econometric Methods, New York: McGraw-Hill, 1960.

Kelly, C. C. and Spilsbury, R. H. Soil Survey of the Okanagan and Similkameen Valley of B. C., Report No. 3 of B. C. Survey. The British Columbia Department of Agriculture in Cooperation with Experimental Farm Service, Dominion Department of Agriculture, 1949.

Keynes, J. M. A Treatise on Probability.

Koutsogianne-Kokkova, A. An Econometric Study of the Leaf Tobacco Market of Greece, Athens, 1962.

Lawrence, H. S. "The Effect of Weather on Agricultural Output": A Look at Methodology, Journal of Farm Economics, Vol. 46, No. 1, February 1964.

Longley, W. V. Some Economic Aspects of the Apple Industry in Nova Scotia, A Thesis for the Degree of Doctor of Philosophy, Nova Scotia Department of Agriculture Bulletin No. 113, 1932.

Mann, H. B. and Wald, A. "On the Statistical Treatment of Linear Stochastic Difference Equations," Econometrica, Vol. 11, 1943.

Snedecor, G. W. and Cochran, W. G. Statistical Method, 6th Ed., 1969.

Tukey, R. B. "Implications of Economics on Orchard Management," The 1969 Apple Forum, Published Proceedings of the First British Columbia Fruit Growers' Association sponsored Horticultural Conference. November 1969.

Van Roechoudt, L. L. Some Factors which Influence the Use of Dwarf and Semi-dwarf Apple Trees for Commercial Orchards in the Okanagan Valley of B. C., Unpublished Master's Thesis, The University of B. C., 1962.

Walpole, R. E. Introduction to Statistics, The MacMillan Co., Collier-MacMillan Limited, London, 1968.

Ware, D. W., Woodward, E. D. and Trevor, H. W. A Study of Apple Production in the Okanagan Valley of British Columbia, Canada Department of Agriculture, Marketing Service - Economic Division Ottawa, January 1952.

Ware, D. W. Organization and Returns of Stone Fruit and Pear Enterprises in the Okanagan Valley, B. C., 1949 - 1950, Department of Agriculture Economic Division, Marketing Service, Ottawa, 1952.

Wonnacott, R. J. and Wonnacott, T. M. Econometrics, John Wiley and Sons, Inc., 1970.

APPENDIX

TABLE I

CLASSIFICATION OF ROOTSTOCK VIGOUR

Size Group	Rootstock	Ultimate Tree Size in Relation to Tree Size on Seedling Roots	Tree Anchorage
Dwarf	M.IX	1/5 to 1/4	Poor
Semi-dwarf	M.26, M.M.106, M.VII, M.IV	1/3	Poor to Fair
Semi-standard	M.II, M.M.III, M.M.104	2/3 to 3/4	Fair to Good
Standard	Seedling, M.XVI, M.XXV, M.M.109	Full size	Good

Source: High-density orchards for B.C. conditions, Research Station, Summerland, B.C., Research Branch, Canada Department of Agriculture. March, 1966.

TABLE II

VARIABLES USED IN MODELS

Variable	Meaning
1	APPLE YIELD PER ACRE
2	DENSITY PER ACRE
3	AGE OF TREE
4	THE AMOUNT OF FERTILIZER APPLIED PER ACRE
5	THE AMOUNT OF SPRAY APPLIED PER ACRE
6	PRUNING AND THINNING LABOUR HOURS SPENT PER ACRE
7	GEOGRAPHICAL DUMMY
8	TREE SIZE INDEX
9	SQUARE OF VARIABLE 2
10	SQUARE OF VARIABLE 3
11	SQUARE OF VARIABLE 4
12	SQUARE OF VARIABLE 5
13	SQUARE OF VARIABLE 6
14	CROSS TERM BETWEEN VARIABLE 2 AND VARIABLE 3
15	CROSS TERM BETWEEN VARIABLE 2 AND VARIABLE 4
16	CROSS TERM BETWEEN VARIABLE 2 AND VARIABLE 5
17	CROSS TERM BETWEEN VARIABLE 2 AND VARIABLE 6
18	CROSS TERM BETWEEN VARIABLE 2 AND VARIABLE 8
19	CROSS TERM BETWEEN VARIABLE 3 AND VARIABLE 4
20	CROSS TERM BETWEEN VARIABLE 3 AND VARIABLE 5
21	CROSS TERM BETWEEN VARIABLE 3 AND VARIABLE 6
22	CROSS TERM BETWEEN VARIABLE 3 AND VARIABLE 8
23	CROSS TERM BETWEEN VARIABLE 4 AND VARIABLE 5
24	CROSS TERM BETWEEN VARIABLE 4 AND VARIABLE 6
25	CROSS TERM BETWEEN VARIABLE 4 AND VARIABLE 8
26	CROSS TERM BETWEEN VARIABLE 5 AND VARIABLE 6
27	CROSS TERM BETWEEN VARIABLE 5 AND VARIABLE 8
28	CROSS TERM BETWEEN VARIABLE 6 AND VARIABLE 8

TABLE III

ESTIMATED SIMPLE LINEAR REGRESSION EQUATION[1]

Equation for the Standard-Tree Group	Significance of b Estimate at .05 Level	r^2
1. VAR 1 = 27750 + 14.15 VAR 2 (6726) (75.43)	NON. SIG.	0.0017
2. VAR 1 = 33950 - 192 VAR 3 (7118) (240)	NON. SIG.	0.0296
3. VAR 1 = 2770 + 126.6 VAR 4 (5360) (310.5)	NON. SIG.	0.0078
4. VAR 1 = 30400 - 22.96 VAR 5 (5603) (68.12)	NON. SIG.	0.0054
5 VAR 1 = 24090 + 60.50 VAR 6 (4496) (41.62)	NON. SIG.	0.0914
Equation for the Semi-Standard-Tree Group		R^2
1. VAR 1 = 15240 - 15.21 VAR 2 (3957) (21.23)	NON. SIG.	0.0085
2. VAR 1 = -3239 + 2114 VAR 3 (4434) (413.5)	SIG.	0.3034
3. VAR 1 = 14810 + 149.1 VAR 4 (2289) (61.55)	SIG.	0.0891
4. VAR 1 = 14010 + 67.60 VAR 5 (3651) (51.18)	SIG.	0.0236
5. VAR 1 = 106.1 + 69.40 VAR 8 (168.9) (219.4)	NON. SIG.	0.0017
Equation for the Semi-Dwarf-Tree Group		r^2
1. VAR 1 = 7688 + 41.73 VAR 2 (5664) (19.65)	SIG.	0.1235
2. VAR 1 = 12520 + 616.1 VAR 3 (7226) (702.2)	NON. SIG.	0.0235
3. VAR 1 = 13110 + 171.1 VAR 4 (2709) (46.07)	SIG.	0.3012
4. VAR 1 = 12330 + 60.21 VAR 5 (2993) (18.07)	SIG.	0.2575
5. VAR 1 = 13620 + 59.62 VAR 6 (3056) (22.11)	SIG.	0.1851
6. VAR 1 = 15260 + 9189 VAR 8 (9345) (26200)	NON. SIG.	0.0038
Equation Total Data		R^2
1. VAR 1 = 18800 + 7.220 VAR 2 (2639) (12.67)	NON. SIG.	0.0028
2. VAR 1 = 13800 + 479.2 VAR 3 (2431) (150.7)	SIG.	0.0795
3. VAR 1 = 16920 + 145.2 VAR 4 (1650) (39.26)	SIG.	0.1047
4. VAR 1 = 16410 + 51.79 VAR 5 (1921) (18.03)	SIG.	0.0659
5. VAR 1 = 15530 + 68.50 VAR 6 (1852) (18.13)	SIG.	0.1087
6. VAR 1 = 341.0 - 245.5 VAR 8 (27.74) (37.81)	SIG.	0.2649

[1]Data in brackets refer to regression coefficient standard errors.

TABLE IV

EVALUATION OF SIZE-CONTROLLING EFFECTS OF ROOTSTOCK, INTERMEDIATE FRAMEWORK STOCK AND STRAIN OF SCION VARIETY ON TOTAL TREE SIZE IN TERMS OF AN INDEX VALUE

		Index Value
A. Standard vigour clonal and seedling roots		1.0
B. Standard vigour framework variety		1.0
C. Standard vigour scion variety		1.0
Reduction in tree size by rootstock in relation to A		
We would rate semi-standard stocks such as	M.II	
	M.M.III	
	A.2	at 0.75
	M.M.104	at 0.85
We would rate semi-dwarf stocks such as	M.M.106	at 0.50
	M.IV	at 0.40
	M.VII	at 0.33
	M.26	at 0.25
We would rate dwarf rootstocks such as	M.IX	at --
	M.VIII	at 0.20
Reduction in tree size by framework stock in relation to B		
We would rate size controlling effect of an intermediate stock such as Haralson (only one on our study)		at 0.75
Reduction in tree size as a result of the use of a spur strain of the scion variety in relation to C		
We would rate reduction of tree size by use of spur type strains		at 0.75

Application

By combining different factors under A, B, and C, a tree-size index value can be established.

For example - Spur Delicious on standard vigour intermediate framework stock on M.IV = 0.74 x 1.0 x 0.40 = 0.30

McIntosh standard vigour intermediate framework stock on M.VIII = 1.0 x 1.0 x 0.33 = 0.33

TABLE VII

INPUT DATA (PER ACRE OF APPLE ENTERPRISE PLOTS)[1]

n = 119

VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR8
Tree-Size Group: Standard						
0.2222E 05	48.00	41.00	7.060	42.38	8.00	1.000
0.2509E 05	48.00	55.00	6.220	112.9	24.00	1.000
4951.	48.00	27.00	20.51	64.74	108.0	1.000
0.1735E 05	48.00	29.00	10.75	65.41	58.00	1.000
0.3[illegible]76E 05	48.00	45.00	14.59	100.0	43.00	1.000
0.2076E 05	48.00	53.00	8.500	58.28	16.00	1.000
0.2383E 05	194.0	8.00	8.900	75.20	21.00	1.000
0.1480E 05	194.0	8.00	8.900	75.20	21.00	1.000
9251.	53.00	18.00	5.710	19.65	17.00	1.000
0.2810E 05	54.00	22.00	21.07	46.44	108.0	1.000
0.2814E 05	73.70	24.00	20.53	114.0	106.0	1.000
0.6222E 05	70.00	22.00	7.460	55.71	116.0	1.000
0.2255E 05	58.00	27.00	22.62	65.06	35.00	1.000
0.3591E 05	108.0	15.00	38.56	24.68	166.0	1.000
0.2835E 05	108.0	12.00	38.56	24.68	166.0	1.000
0.1712E 05	108.0	15.00	6.00	240.0	100.0	1.000
0.5447E 05	108.0	19.00	23.34	74.15	56.00	1.000
0.4013E 05	108.0	17.00	2.540	57.55	105.0	1.000
0.4602E 05	70.00	19.00	17.89	25.59	4.00	1.000
0.4948E 05	48.00	22.00	10.10	29.24	345.0	1.000
0.4602E 05	73.00	28.00	3.690	104.6	34.00	1.000
0.1874E 05	36.00	39.00	5.910	70.03	57.00	1.000
0.1137E 05	48.00	45.00	1.80	1.80	96.00	1.000
Tree-Size Group: Semi-Standard						
4444.	118.0	6.00	8.630	17.84	23.00	0.8500
393.0	118.0	5.00	11.80	18.19	18.00	0.8500
2992.	605.0	9.00	1.380	34.86	52.00	0.7500
1723.	290.0	4.00	7.170	13.66	46.00	0.7500
0.1214E 05	218.0	8.00	18.67	34.22	13.00	0.8500
0.1567E 05	108.0	12.00	15.00	60.00	63.00	0.7500
0.2643E 05	108.0	11.00	17.24	100.7	13.00	0.7500
0.3101E 05	194.0	13.00	45.39	45.71	53.00	0.7500
279.0	194.0	4.00	18.16	5.390	17.00	0.8500
5500.	108.0	6.00	4.730	35.50	17.00	0.7500
0.3456E 05	108.0	11.00	77.22	28.70	107.0	0.8500
0.2621E 05	161.0	14.00	18.83	172.5	52.00	0.7500
0.3155E 05	134.0	14.00	15.57	46.77	43.00	0.7500
0.1210E 05	86.00	11.00	3.720	64.36	240.0	0.7500
8666.	194.0	8.00	1.750	122.2	67.00	0.8500
132.0	132.0	15.00	17.45	40.02	19.00	0.7400
0.2777E 05	132.0	15.00	19.40	153.2	107.0	0.7000
4050.	108.0	6.00	9.570	25.00	34.00	0.7500
0.1561E 05	108.0	7.00	8.640	14.81	86.00	0.7500
4202.	97.00	10.00	16.46	71.88	17.00	0.7500
1071.	218.0	5.00	14.30	40.79	5.00	0.8500
1285.	218.0	5.00	14.47	41.25	6.00	0.8500
0.4597E 05	90.00	22.00	13.85	66.46	72.00	0.7500
0.1224E 05	203.0	13.00	9.560	44.95	31.00	0.7500
0.5530E 05	218.0	14.00	19.46	33.12	1.00	0.7500
0.2858E 05	218.0	11.00	33.64	116.4	2.00	0.7500
0.1619E 05	108.0	12.00	5.780	40.57	54.00	0.7500
0.2565E 05	108.0	13.00	9.720	54.00	92.00	0.8500
0.2370E 05	108.0	19.00	13.12	58.50	160.0	0.8800
0.2032E 05	108.0	13.00	13.12	58.50	87.00	0.7500
0.2044E 05	95.00	22.00	7.620	78.48	25.00	0.7500
3468.	272.0	5.00	5.780	77.14	53.00	0.7500
9346.	97.00	12.00	6.310	49.59	23.00	0.7500
0.1133E 05	129.0	10.00	14.07	47.63	28.00	0.8300
0.7674E 05	218.0	13.00	4.540	10.00	50.00	0.7500
0.5228E 05	605.0	7.00	227.0	145.0	57.00	0.8500
0.1552E 05	104.0	8.00	2.60	99.38	1.00	0.7500
0.1530E 05	108.0	9.00	18.14	22.46	19.00	0.7500
0.4671E 05	108.0	15.00	9.150	58.58	176.0	0.7500
0.3763E 05	186.0	10.00	25.09	31.93	84.00	0.7500
6543.	108.0	5.00	7.330	12.10	32.00	0.8100
4232.	108.0	8.00	6.170	51.43	30.00	0.7500
0.4479E 05	135.0	12.00	12.50	23.98	78.00	0.6700
0.4155E 05	150.0	12.00	13.90	26.63	87.00	0.6100
0.2795E 05	107.0	14.00	8.320	30.10	63.00	0.7800
0.1013E 05	279.0	14.00	103.3	107.9	195.0	0.7500
0.1160E 05	194.0	14.00	71.86	75.05	136.0	0.7500
0.1280E 05	108.0	6.00	17.18	105.2	44.00	0.8400
0.2354E 05	108.0	6.00	17.17	105.2	44.00	0.8400
0.2710# 05	97.00	11.00	5.110	43.47	41.00	0.7500
7642.	218.0	7.00	0.6800	33.07	19.00	0.6500
9559.	218.0	6.00	0.6800	33.07	19.00	0.6700
0.2139E 05	108.0	12.00	20.25	30.66	81.00	0.7500
0.1192E 05	142.0	8.00	30.18	49.91	30.00	0.7500
0.1316E 05	108.0	8.00	9.680	46.80	96.00	0.7500
0.1176E 05	108.0	8.00	9.680	46.80	96.00	0.7500
998.0	108.0	6.00	0.6000E-01	40.00	8.00	0.7500
7347.	151.0	6.00	10.56	73.98	12.00	0.8500
5894.	151.0	6.00	10.56	73.98	12.00	0.6400
2425.	70.00	7.00	5.450	16.36	54.00	0.8400
5955.	134.0	5.00	9.440	19.81	4.00	0.7500
2198.	134.0	5.00	9.440	19.81	4.00	0.7500
Tree-Size Group: Semi-Dwarf						
0.1600E 05	151.0	10.00	18.43	99.90	56.00	0.4000
0.1488E 05	418.0	5.00	20.06	03.35	61.00	0.2500
0.4172E 05	161.0	12.00	13.90	78.66	509.0	0.3300
7495.	218.0	8.00	28.45	62.60	364.0	0.3300
5883.	218.0	8.00	30.04	66.08	82.00	0.3300
4983.	218.0	7.00	18.68	79.41	96.00	0.5000
0.5090E 05	345.0	9.00	271.0	780.0	305.0	0.3100
0.4346E 05	388.0	9.00	50.39	257.0	254.0	0.4000
0.4702E 05	171.0	22.00	16.61	70.35	65.00	0.3300
0.2710E 05	218.0	14.00	95.09	60.21	1.00	0.4000
0.1077# 05	339.0	8.00	57.75	182.0	1.00	0.3300
1995.	97.00	13.00	7.160	21.66	15.00	0.2700
0.1376E 05	453.0	5.00	7.030	75.75	19.00	0.3300
0.2659E 05	453.0	8.00	6.820	76.17	32.00	0.3800
8005.	201.0	10.00	0.1000E-01	82.51	9.00	0.3600
0.1293E 05	97.00	13.00	14.43	27.36	24.00	0.3300
0.6330E 05	605.0	7.00	134.7	217.0	38.00	0.5000
9285.	108.0	13.00	23.25	39.42	37.00	0.3300
777.0	108.0	13.00	23.26	39.54	37.00	0.3300
9934.	108.0	11.00	34.49	85.71	37.00	0.3300
0.2384E 05	108.0	8.00	22.00	127.0	89.00	0.5900
1436.	290.0	18.00	2.870	20.09	34.00	0.5600
1417.	388.0	4.00	3.00	13.32	3.00	0.2500
3182.	142.0	7.00	7.280	104.9	22.00	0.5900
8386.	272.0	5.00	27.20	43.52	14.00	0.3300
0.2161E 05	194.0	7.00	4.430	123.7	12.00	0.3300
0.1128E 05	108.0	9.00	12.73	78.02	30.00	0.3300
0.2890E 05	145.0	9.00	12.61	97.79	81.00	0.3300
Tree-Size Group: Dwarf						
0.1077E 05	339.0	8.00	57.75	182.0	1.00	0.2000
0.2227E 05	134.0	16.00	7.320	41.00	30.00	0.2000
976.0	363.0	5.00	0.1000E-01	54.76	32.00	0.2000
0.2088E 05	388.0	9.00	8.510	39.14	33.00	0.2100
0.2393E 05	387.0	7.00	5.990	1.00	150.0	0.2000
0.2966E 05	388.0	7.00	5.990	1.00	150.0	0.2000

119 Observations Total

118 Degrees of Freedom

[1]See Table II for definition of [illegible]

TABLE VIII

CORRELATION MATRIX FOR SIMPLE LINEAR REGRESSION WITH ONE HUNDRED AND NINETEEN PAIRS OF OBSERVATIONS[1]

Variable	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR8
VAR1	1.0000						
VAR2	0.0526	1.0000					
VAR3	0.2820	-0.4374	1.0000				
VAR4	0.3235	0.4166	-0.0920	1.0000			
VAR5	0.2566	0.2495	-0.0094	0.6882	1.0000		
VAR6	0.3298	0.0072	0.0481	0.2444	0.2745	1.0000	
VAR8	0.1370	-0.5147	0.4314	-0.1310	-0.1726	-0.0669	1.0000

[1]See Table II for definition of variables.

TABLE XI

RESULTS OF EQUALITY OF SLOPE TEST FOR THREE TREE-SIZE GROUPS

Dependent Variable (VAR1) is Apple Yield Per Acre (Quadratic Equation)[1]

Slope Coefficients for Pooled Data

[2]SCP(1)	=	90.190	SCP(2)	=	3430.676
SCP(3)	=	-107.634	SCP(4)	=	408.662
SCP(5)	=	-666.409	SCP(6)	=	88.904
SCP(7)	=	-5393.094	SCP(8)	=	0.021
SCP(9)	=	-19.444	SCP(10)	=	-7.001
SCP(11)	=	-1.511	SCP(12)	=	0.633
SCP(13)	=	7.174	SCP(14)	=	. 1.463
SCP(15)	=	-1.393	SCP(16)	=	1.050
SCP(17)	=	-225.679	SCP(18)	=	-33.096
SCP(19)	=	-4.074	SCP(20)	=	-1.760
SCP(21)	=	-2255.386	SCP(22)	=	5.768
SCP(23)	=	0.214	SCP(24)	=	905.375
SCP(25)	=	0.681	SCP(26)	=	145.471
SCP(27)	=	478.763			

Test for Hypothesis of Slope Coefficients

F	=	0.29
D.F. N_2	=	54
D.F. N_1	=	35
PROB.	=	1.00 (Therefore H_0 is accepted)

[1]See Table II for definition of variables.
[2]SCP stands for Slope Coefficients for pooled data (across tree-size groups).

TABLE XII

SIGNIFICANT COEFFICIENTS AT .05 LEVEL FOR COBB-DOUGLAS MODEL[1]

Dependent Variable (VAR1) is Apple Yield Per Acre

	Constant	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7
	1.8530	0.4932	1.2182	0.2240	0.2533	0.1978	0.1062
ST. Error	(1.3990)	(0.2040)	(0.2218)	(0.0962)	(0.1049)	(0.0746)	(0.0422)
'F' Value		0.0165	0.0000	0.0206	0.0166	0.0089	0.0128

[1]See Table II for definition of variables

TABLE XIII

CORRELATION MATRIX FOR COBB-DOUGLAS MODEL[1]

	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8
VAR1	1.0000							
VAR2	-0.0999	1.0000						
VAR3	0.4406	-0.6386	1.0000					
VAR4	0.3182	0.1268	0.0467	1.0000				
VAR5	0.3140	0.0382	0.1118	0.3783	1.0000			
VAR6	0.3294	-0.1278	0.2321	0.1351	0.0496	1.0000		
VAR7	0.0588	0.0429	-0.1590	-0.1216	-0.1468	-0.0054	1.0000	
VAR8	0.1043	-0.5622	0.3399	-0.0355	0.0390	0.0863	-0.0471	1.0000

[1]See Table II for definition of variables.

TABLE XIV

SIGNIFICANT COEFFICIENTS AT .05 LEVEL FOR QUADRATIC MODEL[1]

Dependent Variable (VAR1) is Apple Yield Per Acre

		ST. Error	'F' Value
Constant	5733.2578	(7318.3136)	
VAR3	2739.7249	(1162.8074)	0.0193
VAR5	398.6005	(85.1116)	0.0000
VAR6	-1096.3665	(192.8729)	0.0000
VAR11	-8.6445	(2.8635)	0.0033
VAR13	0.9406	(0.2208	0.0001
VAR15	3.5579	(1.0176)	0.0008
VAR16	-2.2731	(0.4338)	0.0000
VAR17	1.9925	(0.4324)	0.0000
VAR22	-2671.2339	(1094.0824)	0.0156
VAR23	2.8387	(0.9108)	0.0025
VAR28	866.9939	(143.6294)	0.0000

[1]See Table II for definition of variables.

TABLE XV

CORRELATION MATRIX FOR QUADRATIC MODEL INVOLVING ONLY SIGNIFICANT VARIABLES[1]

Variable	VAR1	VAR2	VAR3	VAR4	VAR5	VAR6	VAR7	VAR8	VAR9	VAR10	VAR11	VAR12	VAR13	VAR14
VAR1	1.0000													
VAR2	0.0526	1.0000												
VAR3	0.2820	-0.4374	1.0000											
VAR4	0.3235	0.4166	-0.0920	1.0000										
VAR5	0.2566	0.2495	-0.0094	0.6882	1.0000									
VAR6	0.3298	0.0072	0.0481	0.2444	0.2745	1.0000								
VAR7	-0.0255	0.0627	-0.2136	-0.0539	-0.1389	-0.1480	1.0000							
VAR8	0.1370	-0.5147	0.4314	-0.1310	-0.1726	-0.0669	-0.0594	1.0000						
VAR9	0.1169	0.9559	-0.2964	0.4453	0.2347	0.0036	0.0687	-0.3863	1.0000					
VAR10	0.1631	-0.3535	0.9545	-0.1004	-0.0072	-0.0242	-0.1942	0.3867	-0.2238	1.0000				
VAR11	0.2945	0.3853	-0.0787	0.9321	0.7473	0.2362	-0.0275	-0.1100	0.4218	-0.0690	1.0000			
VAR12	0.2098	0.1882	-0.0409	0.6780	0.8997	0.2959	-0.1120	-0.1594	0.1595	-0.0355	0.8080	1.0000		
VAR13	0.2410	0.0242	0.0026	0.1890	0.2385	0.9122	-0.1793	-0.1409	0.0042	-0.0369	0.1951	0.2579	1.0000	
VAR14	0.3697	0.6630	0.1975	0.4039	0.2733	0.1194	-0.1019	-0.2557	0.6756	0.1542	0.3413	0.1875	0.0782	1.0000
VAR15	0.3277	0.5443	-0.1236	0.9258	0.5881	0.1536	0.0111	-0.1300	0.6131	-0.1037	0.9012	0.5355	0.1239	0.4471
VAR16	0.2754	0.5247	-0.1565	0.8092	0.9104	0.2475	-0.0896	-0.2855	0.5203	-0.1294	0.8445	0.8621	0.2241	0.4188
VAR17	0.2729	0.4303	-0.1411	0.4837	0.4974	0.7927	-0.0982	-0.3255	0.3872	-0.1456	0.4895	0.5151	0.7150	0.4018
VAR18	0.0952	0.7061	-0.3124	0.4063	0.1047	-0.0807	0.0303	0.1226	0.7370	-0.2613	0.3815	0.0425	-0.0887	0.5163
VAR19	0.3658	0.2117	0.1840	0.9153	0.6274	0.2789	-0.1048	0.0015	0.2514	0.1321	0.8078	0.6170	0.1948	0.4274
VAR20	0.3264	-0.0980	0.5966	0.3901	0.7189	0.1916	-0.2376	0.1537	-0.0325	0.5977	0.4393	0.5903	0.1455	0.2876
VAR21	0.3801	-0.2155	0.4123	0.1095	0.1407	0.8489	-0.1912	0.1959	-0.1514	0.3058	0.1019	0.1506	0.7488	0.1068
VAR22	0.2501	-0.4666	0.9654	-0.1158	-0.0416	0.0178	-0.1814	0.6151	-0.3114	0.9399	-0.0930	-0.0679	-0.0375	0.0854
VAR23	0.2368	0.2507	-0.0579	0.7757	0.8677	0.2872	-0.0909	-0.1612	0.2386	-0.0512	0.8894	0.9827	0.2504	0.2342
VAR24	0.2446	0.2141	-0.0468	0.7816	0.8169	0.4706	-0.0862	-0.1343	0.1882	-0.0581	0.8562	0.9342	0.3980	0.2441
VAR25	0.3274	0.3410	-0.0313	0.8813	0.3845	0.1681	0.0220	0.1148	0.4166	-0.0522	0.7753	0.3319	0.0884	0.3548
VAR26	0.2490	0.1630	-0.0306	0.6509	0.8737	0.5046	-0.1433	-0.1589	0.1297	-0.0428	0.7648	0.9606	0.4554	0.1952
VAR27	0.2884	0.0245	0.2160	0.4741	0.7931	0.1704	-0.0825	0.2728	0.0786	0.1949	0.5103	0.5747	0.0955	0.1865
VAR28	0.3500	-0.2026	0.2055	0.1317	0.0954	0.8251	-0.0897	0.3093	-0.1439	0.0923	0.0972	0.1003	0.6411	0.0043

Variable	VAR15	VAR16	VAR17	VAR18	VAR19	VAR20	VAR21	VAR22	VAR23	VAR24	VAR25	VAR26	VAR27	VAR28
VAR15	1.0000													
VAR16	0.7750	1.0000												
VAR17	0.4254	0.5727	1.0000											
VAR18	0.5597	0.3268	0.2004	1.0000										
VAR19	0.7556	0.6666	0.4265	0.2320	1.0000									
VAR20	0.2985	0.5130	0.2488	-0.1139	0.5352	1.0000								
VAR21	0.0311	0.0673	0.4787	-0.1902	0.2520	0.2899	1.0000							
VAR22	-0.1333	-0.1888	-0.1927	-0.2343	0.1451	0.5609	0.3943	1.0000						
VAR23	0.6532	0.8843	0.5218	0.1296	0.6970	0.5412	0.1404	-0.0844	1.0000					
VAR24	0.6159	0.8238	0.6662	0.1196	0.7480	0.5149	0.2934	-0.0728	0.9527	1.0000				
VAR25	0.8788	0.5187	0.3192	0.5422	0.8180	0.2269	0.1085	-0.0063	0.4600	0.5144	1.0000			
VAR26	0.4986	0.8290	0.6793	0.0236	0.6115	0.5791	0.3215	-0.0641	0.9371	0.9527	0.3272	1.0000		
VAR27	0.4401	0.5922	0.2543	0.1833	0.4918	0.7876	0.1756	0.2594	0.5280	0.4980	0.3914	0.5610	1.0000	
VAR28	0.0490	0.0376	0.4281	-0.0596	0.2338	0.1508	0.9011	0.2432	0.0965	0.2710	0.1943	0.2645	0.1892	1.0000

[1]See Table II for definition of variables.

TABLE XV

CORRELATION MATRIX FOR QUADRATIC MODEL INVOLVING ONLY SIGNIFICANT VARIABLES[1]

TABLE XVI

OBSERVED AND CALCULATED VALUES OF APPLE YIELDS BASED ON QUADRATIC MODEL INVOLVING ONLY SIGNIFICANT VARIABLES

No.	Observed	Calculated	Residual	No.	Observed	Calculated	Residual
1.	22220.	25870.	-3650.0	61.	15304.	12966.	2338.4
2.	25090.	14013.	11077.	62.	46710.	31555.	15155.
3.	4951.0	34297.	-29346.	63.	37630.	19907.	17723.
4.	17350.	31585.	-14235.	64.	6543.0	5165.8	1377.2
5.	36760.	25195.	11565.	65.	4232.0	10588.	-6356.2
6.	20760.	19789.	971.32	66.	44790.	21210.	23580.
7.	23830.	13026.	10804.	67.	41550.	22056.	19494.
8.	14800.	13026.	1774.4	68.	27950.	22411.	5538.9
9.	9251.0	21225.	-11974.	69.	10130.	18988.	-8858.4
10.	28100.	31564.	-3464.2	70.	11600.	25000.	-13400.
11.	28140.	35365.	-7225.2	71.	12800.	10261.	2539.1
12.	62220.	33262.	28958.	72.	23540.	10259.	13281.
13.	22550.	32531.	-9980.8	73.	27100.	15886.	11214.
14.	35910.	26331.	9578.7	74.	7642.0	9317.1	-1675.1
15.	28350.	22101.	6249.2	75.	9559.0	7146.7	2412.3
16.	17120.	27863.	-10743.	76.	21390.	20684.	705.68
17.	54470.	31001.	23469.	77.	11920.	14588.	-2667.8
18.	40130.	29709.	10421.	78.	13160.	15346.	-2185.6
19.	46020.	24938.	21082.	79.	11760.	15346.	-3585.6
20.	49480.	46341.	3139.2	80.	998.0	4169.1	-3171.1
21.	46020.	30649.	15371.	81.	7347.0	7592.3	-245.27
22.	18740.	28916.	-10176.	82.	5894.0	7115.3	-1221.3
23.	11370.	29489.	-18119.	83.	2425.0	9676.1	-7251.1
24.	4444.0	6903.5	-2459.5	84.	5955.0	3866.3	2088.7
25.	393.0	5115.1	-4722.1	85.	2198.0	3866.3	-1668.3
26.	2992.0	26570.	-23578.	86.	16000.	17459.	-1458.8
27.	1723.0	6211.3	-4488.3	87.	14880.	11282.	3597.9
28.	12140.	14213.	-2072.7	88.	41720.	45607.	-3887.1
29.	15670.	20162.	-4492.3	89.	7495.0	20675.	-13180.
30.	26430.	17032.	9398.0	90.	5883.0	14922.	-9039.5
31.	31010.	26450.	4560.5	91.	4983.0	15094.	-10111.
32.	279.0	4730.9	-4451.9	92.	50900.	50890.	9.5859
33.	5500.0	5677.7	-177.65	93.	43460.	23871.	19589.
34.	34560.	17805.	16755.	94.	47020.	34947.	12073.
35.	26210.	26874.	-664.44	95.	27100.	34165.	-7065.5
36.	31550.	22932.	8617.7	96.	10770.	25954.	-15184.
37.	12100.	31349.	-19249.	97.	1995.0	16300.	-14305.
38.	8666.0	15275.	-6608.6	98.	13760.	8904.6	4855.4
39.	132.0	23189.	-23057.	99.	26590.	18478.	8112.4
40.	27770.	28988.	-1217.9	100.	8005.0	13991.	-5986.5
41.	4050.0	7442.4	-3392.4	101.	12930.	17348.	-4417.9
42.	15610.	12657.	2952.9	102.	63300.	47204.	16096.
43.	4202.0	14711.	-10509.	103.	9285.0	18777.	-9491.5
44.	1071.0	6545.6	-5474.6	104.	777.0	18779.	-18002.
45.	1285.0	6640.8	-5355.8	105.	9934.0	17388	-7453.7
46.	45970.	31403.	14567.	106.	23840.	15579.	8261.2
47.	12240.	22657.	-10417.	107.	1436.0	34242.	-32806.
48.	55300.	24749.	30551.	108.	1417.0	3261.1	-1844.1
49.	28580.	24561.	4018.6	109.	3182.0	9171.4	-5989.4
50.	16190.	18802.	-2611.8	110.	8386.0	7242.5	1143.5
51.	25650.	23354.	2296.5	111.	21610.	8799.6	12810.
52.	23700.	34929.	-11229.	112.	11280.	12355.	-1075.3
53.	20320.	22872.	-2551.9	113.	28900.	16522.	12378.
54.	20442.	28198.	-7755.9	114.	10770.	24339.	-13569.
55.	3468.	9224.4	-5756.4	115.	22270.	23062.	-792.12
56.	9346.0	16222.	-6876.1	116.	976.0	8091.6	-7115.6
57.	11330.	15743.	-4412.6	117.	20880.	19102.	1778.1
58.	76740.	23925.	52815.	118.	23930.	22415.	1515.1
59.	52280.	60396.	-8115.6	119.	29660.	22435.	7224.7
60.	15520.	7730.5	7789.5				

Autocorrelation Coefficient 0.024

Durbin-Watson D Statistic 1.948

INFORMATION SHEETS

(COMPLETE SET OF APPLE-ENTERPRISE-DATA SHEETS USED IN PART FOR THESIS STUDY[1])

[1]Designed by the Economics Branch - Vancouver, Canada Department of Agriculture, B.C.

ENTERPRISE TREE FRUIT RECORD
Economics Branch, Canada Department of Agriculture
6660 N. W. Marine Drive, Vancouver 8, B. C.

Study Year

Name:	Record No.
P.O. Address:	Date Taken:
District:	Taken by:
Marketing Point:	Check by:

Land Description and Value

	Orchard		Other Improved		Unimproved		Waste	Total Acres
	Acres	Value/Acre	Acres	Value/Acre	Acres	Value/Acre	Acres	
Owned								
Rented								

Additional Acres Owned Land Suitable for Orchard

LAND IMPROVEMENTS

	Clearing	Breaking	Disc and Harrow	Picking Roots Stones
Acres				
Cover description				
Hours - Farm tractor				
- Unpaid Labor				
Cost - Hired tractor				
- Hired labor				
Other costs				
Total Costs				

LAND PURCHASES & SALES

	Purchases	Sales	
Price			
Paid			
Received			
Owing			
		Yes	No
Purchased Land Cropped			
Sold Land Cropped			
Date of Purchase or Sale			

	Orchard	Other improved	Unimproved
Acres Purchased			
Acres Sold			

Expenditures on New Construction, Improvements and Repairs Associated with Land

	CASH COSTS							Value of Farm Contributions					Total	Cash Cost
	Hired Mach.		Mater'l	Hired Labor		Total		Farm Tractor		Mater'l	Unpaid Labor		Value	Repairs
	Hrs.	Costs	Costs	Hrs.	Costs	Cash		Hrs.	Cost	Value	Hrs.	Value		
Irrg. Ditches														
Drainage														
Farm Roads														
Fences														
Other														

Inventory of Buildings

	Year Built	Replacement Cost Beg. of Year	Type of Cons't	Cash Cost New Bldg. & Imp.: Equip & Material	Cash Cost New Bldg. & Imp.: Hired Labor	Value of Farm Contributions: Farm Tractor Hrs.	Value of Farm Contributions: Farm Tractor Value	Value of Farm Contributions: Matr'l Value	Value of Farm Contributions: Unpaid Labor Hrs.	Value of Farm Contributions: Unpaid Labor Value	Total Cost	Repairs
Machine Shed												
Tool Shed												
Garage												
Pump House												
Pickers Cabins												
Road Side Stand												

STANDARD APPLES

TREE FRUIT INVENTORY

Variety	Spacing	Trees/ Acre	Planted 1969	1 to 5 Yrs.	6 to 10 Yrs.	11 to 20 Yrs.	Over 20 Yrs.	Total Trees	Est. Acres	Inter-Plants	Winter 69 Killed	Winter 69 Dam.
Red Delicious												
Golden Delicious												
McIntosh												
Winesap												
Newton												
Rome Beauty												
Spartan												
Stayman												
Tydeman's Red												

Variety	Spacing	Trees/ Acre	Planted 1969	Planted 1968	Planted 1967	Planted 1966	Planted 1965-61	Over 5 Yrs.	Total Trees	Est. Acres	Winter 1969	
											Killed	Dam.
Semi-Standard Apples												
Semi-Dwarf Apples												
Dwarf Apples												

Variety	Spacing	Trees/ Acre	Planted 1969	1 to 5 Yrs.	6 to 10 Yrs.	11 to 20 Yrs.	Over 20 Yrs.	Total Trees	Est. Acres	Winter 1969	
										Killed	Damaged
Pears											
Peaches											

Variety	Spacing	Trees/ Acre	Planted 1969	1 to 5 Yrs.	6 to 10 Yrs.	11 to 20 Yrs.	Over 20 Yrs.	Total Trees	Est. Acres	Winter 1969	
										Killed	Damaged
Apricots											
Cherries											
Plums & Prunes											

PHYSICAL DATA - RE SPECIFIC APPLE PLOTS

		Plot of Intensive planting	Plot of Standard Planting
Variety			
Main Root Stock			
Size of Block Ac.			
Spacing			
Total No. of Trees			
Trees per Acre			
Year Planted			
Winter Damage 69			
Trees Killed No.			
Trees Damaged No.			
Air Drainage	Good		
	Fair		
	Poor		
Direction of Slope			
Degree of Slope	Level		
	Slight		
	Moderate		
	Steep		
Soil Type			
Cover Crop	Yes		
	No		

PRODUCTION AND RECEIPTS FROM SPECIFIC APPLE PLOTS

	Intensive	Standard
Variety		
Type of Pack		
Extra Fancy: Large		
Medium		
Small		
Fancy: Large		
Medium		
Small		
Cee		
Culls		
Receipts		
Extra Fancy		
Fancy		
Cee		
Rebates		
Cull Returns		
Total Receipts		
Farm Sales: Lbs.		
Receipts		
Home Use: Lbs.		
Value		

GRADE AND SIZE OF FRUIT DELIVERED TO THE PACKING HOUSE

Name of Packing House -

APPLES

Variety	Type of Pack	Extra Fancy			Fancy			Cees	Culls
		Large	Medium	Small	Large	Medium	Small		

Apple Receipts

Variety	Type of Pack	Extra Fancy	Fancy	Cee	Rebates	Cull Return	Total Receipts	Cull Charge

GRADE AND SIZE OF FRUIT DELIVERED TO THE PACKING HOUSE

P E A R S

Variety	Type of Pack	Extra Fancy			Fancy			Cees	Culls
		Large	Medium	Small	Large	Medium	Small		

Pear Receipts

Variety	Type of Pack	Extra Fancy	Fancy	Cee	Rebates	Total Receipts	Cull Charge

	GRADE AND SIZE				RECEIPTS					
Plums and Prunes Variety	Select	No. 1	No. 2	Culls	Select	No. 1	No. 2	Rebates	Total Receipts	Cull Charge

Peaches Variety	Domestic Grade		Culls	Domestic Grade		Rebates	Total Receipts	Cull Charge
	No. 1	No. 2		No. 1	No. 2			

Apricots Variety	Domestic Grade			Culls	Domestic Grade			Rebates	Total Receipts	Cull Charge
	No. 1	No. 2	No. 3		No. 1	No. 2	No. 3			

Cherries Variety	No. 1	Orchard Run	Culls	No. 1	Orchard Run	Rebates	Total Receipts	Cull Charge

FRUIT SOLD AND USED ON FARM

	Farm Sales		Used on Farm			Farm Sales		Used on Farm	
	Pounds	Receipts	Pounds	Value		Pounds	Receipts	Pounds	Value
Apples					Peaches				
					Apricots				
Pears									
					Cherries				
Plums and Prunes									
					Totals	xxx		xxx	

FERTILIZER USED

Kind of Fertilizer	Total Orchard		Intensive Apple Plot			Standard Apples			Other Specified Fruit	
	Quantity	Cost	Lbs/Tree	Quantity	Costs	Lbs/Tree	Quantity	Costs	Quan.	Costs

	SPRAY MATERIAL							
Kind of Spray	Total Orchard		Intensive Apple Plot		Standard Apple Plot		Other Specific Fruit	
	Quantity	Costs	Quantity	Costs	Quantity	Costs	Quantity	Costs
Boron								
Zinc								
Magnesium								
Manganese								
Iron								
Urea								
Dinitrocresol								
Naphthalene acetmide								
(Amid Thin)								
Sevin								
Triethanolamine Salt								
of 2, 4, 5 - T.P.								
Naphthalene Acetic								
Acid (N.A.A.)								
Alar								
Dormant Spray								
Morastan								
Kelthane								
Tedion								
Ethion								
Kava thane								
Movocide								
Dimethoate								
(Cygon, Roger)								
Total	xxx		xxx		xxx		xxx	

Kind of Spray	Total Orchard		Intensive Apple Plot		Standard Apple Plot		Other Specific Fruit	
	Quantity	Costs	Quantity	Costs	Quantity	Costs	Quantity	Costs
Guthion								
Perthane								
Supreme and Superior Type Oils								
D.D.T.								
Parathion								
Diazinion								
Thiodan								
Para dichlorobenzene								
Lime Sulphur								
Glyodin-Dodine (Glyodax)								
Dodine (Cyprex)								
Dichlone (Phygon)								
Ferbam								
Maneb								
Zirim								
Captan								
Bordeaux								
Botran								
Malathion								
Wettable Sulphur								
Paste Sulphur								
Fixed Copper								

COST OF SEEDS AND PLANTS

	Total Orchard		Intensive Apple Plot		Standard Apple Plot		Other Specific Fruit	
	Quantity	Cost	Quantity	Cost	Quantity	Cost	Quantity	Cost
Fruit Trees								
Grass and Plants								
Total								

CUSTOM WORK

	Rate	Total Orchard		Paid Intensive Apple Plot	Paid Standard Apple Plot	Paid Other Specific Fruit
		Received	Paid			
Plowing						
Discing						
Mowing						
Raking						
Ditching						
Spraying						
Hauling						
Other trucking						
Total	xxx					

GENERAL MACHINERY AND EQUIPMENT

	Total Orchard					Share to Intensive Apple Plot	Share to Standard Apple Plot	Share to Other Specific Fruit
	No.	Begin of Yr. Value	Purchases	Sales	Cost of Repairs			
Irrigation -								
Orchard								
Spray Costumes & Mask								
Pruning Equipment								
Props								
Ladders								
Picking Bags								
Orchard Boxes								
Equipment - Pickers Cabin								
Ditcher								
Sub Total	xxx							

	Total Orchard					Share to Intensive Apple Plot	Share to Standard Apple Plot	Share to Other Specific Fruit
	No.	Beg. of Year Value	Purchases	Sales	Cost of Repairs			
Plow								
Disk								
Row Crop or Field Cultivator								
Harrows								
Mower								
Rake								
Hand Sprayer								
Trailer								
Wagon								
Chain Saw								
Electric Motors								
Small Tools and								
Garden Tools								
Total								

	Car	Truck	Tractor	Garden Tractor	Sprayer	Giraffe Squirrel etc.	Roto Mower	
Year								
Make								
Size								
Miles for Year								
Miles to Farm								
Hours Used								
Value Beg. of Yr.								
Purchase Price								
Sales Price								
Operating Costs								
Cost of Fuel								
Oil and Grease								
Repairs								
Tires								
Licence								
Insurance								
Total Operating Costs								
Proportion of Totals to -								
Intensive Apple Plot								
Standard Apple Plot								
Other Specific Fruit								

N. B. If gross figures only available on costs estimate cost of car operation or truck if used instead of car.

	Rate of Wages	Total Orchard			Intensive Apple Plot			Standard Apple Plot			Other Specific Fruit		
		Total Hours	Total Wages	Board	Total Hours	Total Wage	Board	Total Hours	Total Wage	Board	Total Hours	Total Wage	Board
Hired: Month													
Day													
Piece Work													
Family - Daughter													
Son (Age)													
Son (Age)													
Wife													
Operator													

LABOR INPUTS RE PLOTS

	Intensive Apple Plot			Standard Apple Plot		
	Hired	Family	Operator	Hired	Family	Operator
Pruning, Grafting, Repairing and Removing trees						
Cultivating						
Mowing						
Fertilization						
Spraying						
Thinning, propping						
Irrigating						
Picking						
Distributing & hauling boxes to packing house						
Collecting and storing boxes						

FARM LIABILITIES

Borrowed From	Purpose	Amt. Owing Beg. of Yr.	Borrowed During the Year			Paid During Year		Owing End of Year
			Amt.	Term	Rate	Princ.	Int.	
Provincial Gov't								
Farm Credit								
Farm Improvement								
V.L.A.								
Bank								
Credit Union								
Mortgage Co.								
Finance Co.								
Machinery Co.								
Other								
Current Borrowing								
Bank								
Credit Union								
Other								

RECEIPTS

Current Receipts	Total Orchard	Intensive Apple Plot	Standard Apple Plot	Other Specific Fruit
Fruit: Apples				
Pears				
Plums and Prunes				
Peaches				
Apricots				
Cherries				
Farm Sales				
Custom Work				
Non-Farm Earnings				
Other				
Total -				

Capital Receipts				
Real Estate Sales				
Power Equipment Sales				
General Equipment Sales				
Other				
Total -				

MEMBERS OF FAMILY

	Sex	Age	Months at Home
Operator			
Wife		xxx	
Children: 1			
2			
3			
4			
5			
6			
7			
8			
9			
Others			

Year Operator Started on this Farm	
Acres Orchard	
Acres Improved	
Acres Unimproved	

EXPENSES

	Total Orchard	Intensive Apple Plot	Standard Apple Plot	Other Specific Fruit		Total Orchard	Intensive Apple Plot	Standard Apple Plot	Other Specific Fruit
Current Expenses					Current Expenses Con'd				
Cash Rent					Electric(½ to farm)				
Land Taxes					Phone(½ to farm)				
Irrigation- Water Tax					Freight & Express				
Water Toll					Accounting				
Electricity					Interest(current)				
Gas and Oil					Membership Fees				
Fire Insurance					Orchard Box Rental				
Hail Insurance					Other				
Repairs - Land									
Buildings									
Cull Charges									
Plants & Seeds Purchased									
Fertilizer									
Spray Material					Total Current				
Operating Costs of									
all Equipment					Capital Expenses				
Labor: Wages									
Un. Ins.					New Cons't Land				
C.P.P.					Buildings				
Liab. Ins.					Land Improvements				
Custom Work					Power Equip. Pur.				
Weed Sprays					General Equip. Pur.				
Small Hardware					Other				
Misc. Oil & Grease					Total Capital				

LABOR TIME SHEET

Date	Description of Work	Intensive Apple Plot				Semi-Standard or Standard Apple Plot			
		Full-time Employee and Operator	Family Labor		Casual Hired	Full-time Employee and Operator	Family Labor		Casual Hired
			Adult	Under 15 Yrs.			Adult	Under 15 Yrs	
		Hours				Hours			

CASH EXPENSES

Date	Description (including custom work)	Intensive Apple Plot		Standard or Semi-Standard Apple Plot	
		Quantity or Hrs.	Cost	Quantity or Hrs.	Cost
				Y	

CASH RECEIPTS

Date	Kind of Apple	Intensive Apple Plot					Semi-Standard or Standard Apple Plot				
		Pounds	Grade	Culls Pounds	Total Receipts	Cull Charges	Pounds	Grade	Culls Pounds	Total Receipts	Cull Charges

NET ESTABLISHMENT COST IN 1969 OF AN INTENSIVE APPLE PLOT PLANTED IN 19__

Kind of root stock	Land value per acre
No. of trees in plot	Value of irrigation system
Spacing	Cost of trees
Area planted - acres	Est. value of equipment used
Receipts from crop sales	Est. value of equipment chargeable to plot
	Est. operating cost of equipment to plot
Total Orchard Size - Acres	Est. of taxes

Date	Quantity	Hours 1/	Description of Item	Custom Work	Fertilizer	Irrigation	Pruning Thinning	Spraying	Tree Replacement	Weed Control	Mowing	Picking & Hauling	Sundry
				Dollars									

1/ If operator or family labor - Do not put in value but indicate item applicable with "✓".

NET ESTABLISHMENT COST IN 1969 OF AN INTENSIVE APPLE PLOT PLANTED IN ________ Cont'd

Date	Quantity	Hours 1/	Description of Item	Custom Work	Fer-tilizer	Irriga-tion	Pruning Thin-ing	Spray-ing	Tree Replace-ment	Weed Con-trol	Mow-ing	Pick-ing & Hauling	Sundry
				Dollars									

1/ If operator or family labor - Do not put in value but indicate item applicable with "✓".