

5105  
E55  
no. 381  
cop. 2

# A Decision Model for Apple Harvester Selection



Special Report 381, April 1973

Agricultural Experiment Station, Oregon State University

## A DECISION MODEL FOR APPLE HARVESTER SELECTION

### Authors

T. H. Rudkin  
R. D. Langmo  
A. N. Halter

T. H. Rudkin, former graduate research assistant, Department of Agricultural Economics, Oregon State University. R. D. Langmo, Associate Professor and Industrial Engineer, Department of Agricultural Economics, Oregon State University. A. N. Halter, Professor, Department of Agricultural Economics, Department of Statistics, Oregon State University. This report is based on a Master of Sciences thesis, "A Sampling Procedure to Detect Alternative Actions in a Decision Model for Choosing Among Apple Harvesting Systems."

A. G. Berlage, Agricultural Engineer, USDA, Tree Fruit Research Center, Wenatchee, Washington, and W. M. Mellenthin, Superintendent of Mid-Columbia Experiment Station, Hood River, Oregon, are acknowledged for helpful suggestions and primary data.

## SUMMARY

Numerous analytical aids are available to assist in the decision problems of choosing among alternative harvesting systems. Many of these techniques fail to recognize the effect of apple orchard yield variability on system performance and economic efficiency. Also, economic principles essential to decision making are not always conscientiously applied. Thus, the objective of this study was to formulate a performance test procedure for alternative apple harvesting systems. In pursuit of this objective a decision model was formulated which recognizes (1) the natural condition of yield variability among apple trees in an orchard, (2) the variability due to harvesting system productivity, and (3) costs as they are related to system productivity and economic life. Necessary simulation program and statistical test information was obtained from actual orchard yields and harvesting time data.

The prescriptive character of the work precluded findings of a conclusive nature. However, sensitivity tests conducted on the present value of total cost computer simulation programs indicated that implementation of the developed guidelines will clarify factors that influence apple harvester performance, and aid the decision making process of producers.

## CONTENTS

Introduction .....	1
Framework and Application of Decision Theory .....	3
System Productivity and Fruit Density .....	25
Calculations of Payoffs in the Decision Model and Selection of an Action .....	37
Bibliography .....	47
Appendix .....	49

## INTRODUCTION

Mechanization in general has brought many rewards to the agricultural sector of the U.S. economy. Apart from its effect of increasing labor productivity by serving as a substitute for human efforts, it has greatly reduced the drudgery related to many farming operations. Also, mechanization has introduced farming techniques which have increased manyfold the resources man can manage.

In agricultural harvest operations, the development and refinement of mechanical harvesting systems have progressed most rapidly for row and field crops. For specialty crops, particularly tree fruits whose quality and appearance are easily damaged by harsh mechanical operations, mechanization has progressed slowly. For some fruits, such as apples, freestone peaches, and pears, it still takes as long to pick an acre of trees as it did 30 years ago (Simpson, 1966).

Several factors have impeded development of harvesting aids in the tree fruit industry:

- 1) Fear of doing lasting damage to a tree's trunk, roots, limbs, and foliage by harsh mechanical actions.
- 2) Availability of adequate seasonal labor for harvest at relatively low cost.
- 3) Acceptance by producers and packers to fruit damage from hand picking.
- 4) Limitations of orchard environment on mechanical devices.
- 5) Complexities of mechanical design necessary because tree fruits do not ripen uniformly.

Many ingenious mechanical aids to replace or assist manual apple harvesting have been proposed, designed, refined, and manufactured. Some are in commercial production; others are feasible from engineering standpoints, but await further improvements to make them economically practicable. Generally, mechanical apple harvesting systems may be grouped into four categories: (1) tree shakers, (2) air or water jet streams, (3) multiman platforms and towers, and (4) man-positioners.

#### Choosing Among Harvesting Systems

For an apple producer, the problem of choosing between mechanical harvesting and continued use of conventional hand labor is basically the same as that commonly faced by other farmers, i.e., achieving a feasible and proper balance between the level of mechanization and the constraints imposed by the availability of his land, labor and capital resources. He is chiefly interested in improving his harvest operation by decreasing his labor requirement, and thereby labor cost. However, because of generally higher fixed costs of mechanical harvesting systems, a mechanical system to be competitive with harvesting costs, may have to show a substantial increase in productivity.

The ease with which a producer carries out a mechanization development program depends on the availability of adequate information on mechanical system design, capability, economic efficiency and the personal capacity to decide which system best meets his particular needs and resource conditions. Assuming an apple producer has adequate information about the capabilities of available mechanical harvesters, and thereby has eliminated those not suited to his apple production enterprise, he needs an effective and valid analytical procedure to measure costs and economic efficiencies. The farm advisory literature, both public and private, suggests numerous procedures to aid apple producers with this process. A majority of these procedures can be placed in one of the following categories: (1) time study, (2) break-even analysis, and (3) cost study. A review of these analytical aids shows they fail to give consideration to the natural yield variability of apple production. Yield variability has been given substantial empirical confirmation yet these procedures do not suggest a means for measuring this variability or utilizing it in a practical decision situation.

For an apple producer to economically evaluate alternative apple harvesting systems, he must have a decision model that will take into account: (1) the variability of yields among apple trees, (2) the variability in productivity among apple harvesting systems, and (3) the fixed and variable costs as they are related to the productivity and economic life of the harvesting system.<sup>1/</sup> It is the purpose of this report to present such a decision model and to provide research agencies concerned with the development and performance testing of harvesting systems a guide for the collection of performance data.

#### FRAMEWORK AND APPLICATION OF DECISION THEORY

Apple producers face the task of carefully evaluating the alternatives and consequences of operating decisions which have a significant effect on the economic success of the firm. In many situations, the decision making process appears to rely more on subjective judgments, intuition, and rules of thumb than on any systematic procedure based on economic principles. Modern decision theory seeks to provide a formal, systematic approach to analyzing decision alternatives when knowledge about reality is imperfect.

In the decision process, a decision maker usually has several actions available in his attempt to maximize expected utility. The outcomes of alternative actions depend on certain conditions which cannot be known with certainty. These conditions are called states of nature. A lack of certainty about states of nature is characterized by a probability distribution over the states of nature. The combination of a particular action and occurrence of a specific state of nature will result in a specific state of affairs or consequences for a decision maker. These consequences yield a decision maker costs and returns which in turn provide a level of satisfaction or utility. The monetary outcomes or the level of utility received from a given action and state of nature combination are called the payoff to a decision maker. A measure of utility would be necessary to convert monetary outcomes to utility payoffs.

---

<sup>1/</sup> Assuming equivalent marketing conditions.

## Example of a Harvesting System Decision Problem

Suppose an apple producer is contemplating purchase of one of several apple harvesting systems. He knows that for a mechanical harvesting system to be an economically feasible alternative to his hand labor and ladder method, it must increase the productivity of the harvest operation. He knows that most mechanical harvesting systems demand higher fixed investments than do conventional methods. Furthermore, the producer believes variable cost for a harvesting system is related directly to the amount of time it takes to pick his orchard. The amount of time required depends on the number of trees in his orchard and on the size and number of fruit on the trees, i.e., the density of fruit. Finally, since harvesting equipment usually has an extended useful life beyond one harvest season, the producer realizes he will incur a flow of costs throughout its useful life.

Suppose the apple producer is considering three possible actions:

$A_1$  = purchase a man-positioner aid,

$A_2$  = purchase a multiman platform system,

$A_3$  = use conventional hand labor and ladder method.

His choice among these actions depends on three states of nature:

$\theta_1$  = fruit density is greater than 8.0 trees per bin,

$\theta_2$  = fruit density is between 4.0 and 7.9 trees per bin,

$\theta_3$  = fruit density is between 0.1 and 3.9 trees per bin.

The decision maker assigns (estimates) a probability to occurrence of each state of nature:

$$P(\theta_1) = 0.20,$$

$$P(\theta_2) = 0.50,$$

$$P(\theta_3) = 0.30.$$

Table 1 shows the hypothetical present value of cost per bin incurred by choosing a particular action given the occurrence of a specific state of nature.

The decision maker's objective could be minimization of cost or minimization of disutility for cost.<sup>2/</sup> Either criterion involves calculation of expected value for each action. Expected costs per bin for each action are calculated as follows:

$$E(A_1) = P(\theta_1) C(\theta_1, A_1) + P(\theta_2) C(\theta_2, A_1) + P(\theta_3) C(\theta_3, A_1);$$

$$E(A_2) = P(\theta_1) C(\theta_1, A_2) + P(\theta_2) C(\theta_2, A_2) + P(\theta_3) C(\theta_3, A_2);$$

$$E(A_3) = P(\theta_1) C(\theta_1, A_3) + P(\theta_2) C(\theta_2, A_3) + P(\theta_3) C(\theta_3, A_3).$$

The expected costs per bin are \$3.60, \$3.65, and \$3.50 for actions  $A_1$ ,  $A_2$ , and  $A_3$ , respectively. The decision maker would choose that action which has the smallest expected cost per bin if his decision criterion was minimization of cost. He would choose action  $A_3$ , the conventional hand labor and ladder method.

To apply the criterion of minimization of disutility not only must the mean (expected value) be calculated, but also the variance of each action.

Variance of cost per bin for each action is calculated as follows:

$$V(A_1) = P(\theta_1) [C(\theta_1, A_1) - E(A_1)]^2 + P(\theta_2) [C(\theta_2, A_1) - E(A_1)]^2 + P(\theta_3) [C(\theta_3, A_1) - E(A_1)]^2;$$

$$V(A_2) = P(\theta_1) [C(\theta_1, A_2) - E(A_2)]^2 + P(\theta_2) [C(\theta_2, A_2) - E(A_2)]^2 + P(\theta_3) [C(\theta_3, A_2) - E(A_2)]^2;$$

$$V(A_3) = P(\theta_1) [C(\theta_1, A_3) - E(A_3)]^2 + P(\theta_2) [C(\theta_2, A_3) - E(A_3)]^2 + P(\theta_3) [C(\theta_3, A_3) - E(A_3)]^2.$$

<sup>2/</sup> A criterion of minimization of monetary cost may be identical to minimization of disutility for cost. That is, utility functions can take different forms for different individuals, each form possibly yielding a different solution. Thus, minimization of expected disutility for cost can yield a different result than minimization of expected cost but the same solution if a linear utility function is assumed. The difference lies in the shape of each decision maker's utility functions.



Table 1. Hypothetical present values of cost per bin,  $C(\Theta, A)$  for making a decision among alternative harvesting systems

State of nature	Action			Probability of state of nature
Fruit density in trees per bin	$A_1$ : Man-positioner system	$A_2$ : Multiman platform system	$A_3$ : Hand and ladder system	$P(\Theta)$
$\Theta_1$ : greater than 8.0	$C(\Theta_1, A_1) = \$5.00$	$C(\Theta_1, A_2) = \$6.00$	$C(\Theta_1, A_3) = \$8.00$	0.20
$\Theta_2$ : 4.0 to 7.9	$C(\Theta_2, A_1) = \$4.00$	$C(\Theta_2, A_2) = \$4.00$	$C(\Theta_2, A_3) = \$3.20$	0.50
$\Theta_3$ : 0.1 to 3.9	$C(\Theta_3, A_1) = \$2.00$	$C(\Theta_3, A_2) = \$1.50$	$C(\Theta_3, A_3) = \$1.00$	0.30
E (A): Expected cost	\$3.60	\$3.65	\$3.50	
V (A): Cost variance	\$1.24	\$2.55	\$5.97	

The variances for the decision problem described in Table 1 are 1.24, 2.55, and 5.97 for actions  $A_1$ ,  $A_2$ , and  $A_3$ , respectively. The relationship between expected utility and the moments (mean and variance) of the probability distribution over the states of nature are shown in Figure 1.

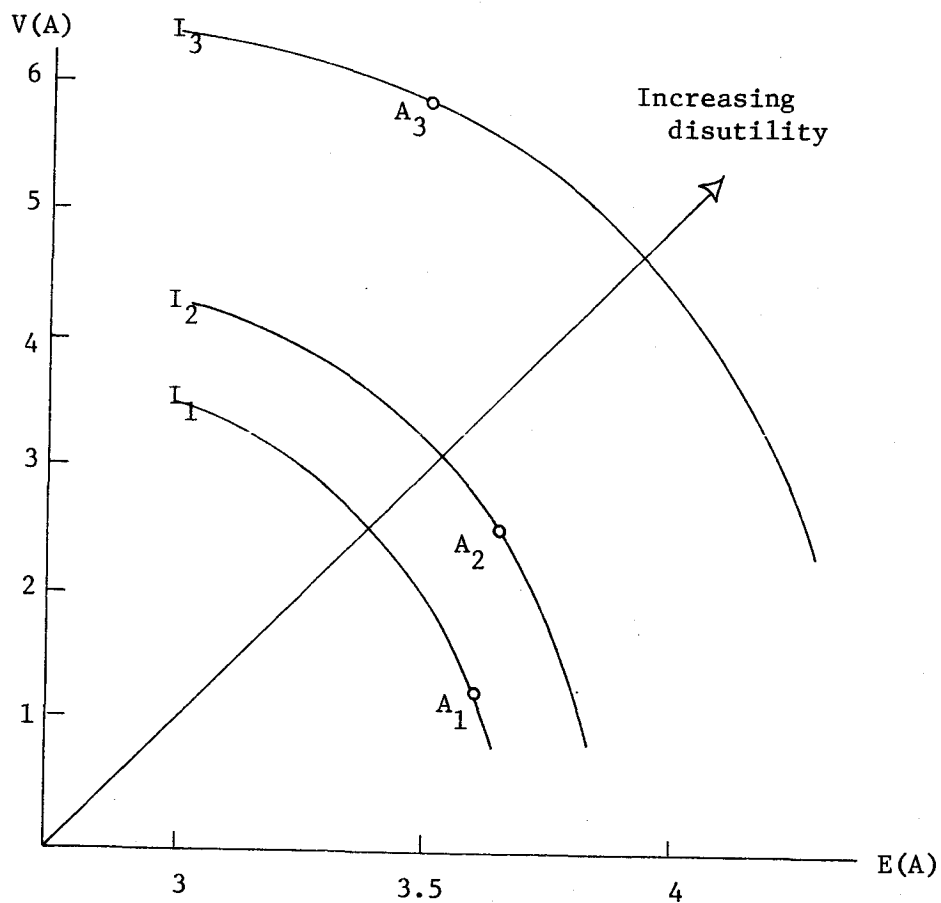


Figure 1. Hypothetical indifference curves and the three actions.

The mean ( $E(A)$ ) of cost per bin is measured along the horizontal axis, and variance ( $V(A)$ ) along the vertical axis. Utility levels are shown by indifference curves joining combinations of  $E(A)$  and  $V(A)$  to which the decision maker is indifferent. The set of indifference curves in Figure 1 is referred to as gambler's indifference curves or E-V curves.<sup>3/</sup> Because Figure 1 shows levels of disutility for cost, the indifference curves are labeled from left to right in order of increasing disutility ( $I_1 < I_2 < I_3$ ).

If an additional action  $A_x$  had been among those available, this particular individual would be indifferent between  $A_x$  and action  $A_3$ . However, he would prefer action  $A_2$  to either  $A_3$  or  $A_x$  because indifference curve  $I_2$  represents less disutility than does  $I_3$ . Action  $A_1$  would be chosen since it is on a lower indifference curve and provides the minimum expected disutility. Thus a criterion of minimization of disutility may result in selection of a different action ( $A_1$ ) over that chosen by minimization of cost ( $A_3$ ) depending upon the shape of the gambler's indifference surface. The importance of calculating the variance of each action, in addition to the mean when the decision maker's utility function is nonlinear is illustrated by the shape of the indifference surface of Figure 1.

#### Expected Value and Variance for Nondiscrete States of Nature

The state of nature for the decision problem in this study, described in detail in subsequent sections, is characterized as apple orchard fruit density. As in many decision problems, states of nature cannot be easily categorized into discrete units or increments.

---

<sup>3/</sup> For further discussion of E-V curves, see Halter and Dean (1971), Chapter IV.

Expected values and variances for alternative actions must be found over a continuous distribution of states of nature. Given that mean ( $E(A)$ ) and variance ( $V(A)$ ) values have been calculated over a continuous distribution, application of a decision criterion for solution of a decision problem can proceed analogously to that of a discrete case. That is, in a case of minimization of expected cost, the minimum  $E(A)$  is chosen; in a case of minimization of expected disutility, the E-V analysis is applied.

### Components of the Harvesting System Decision Problem

An association has been made between the theoretical framework and practical application of decision theory by introducing, through the example, the components describing a decision model formulation. The following sections define in detail these components of the decision model.

#### The Alternative Actions

For harvesting systems to be alternative decision making actions, they necessarily must have economically important distinguishing characteristics. Useful economic life and system productivity are two major characteristics which play a substantial role in the economic decision problem and are used here as distinguishing factors among harvesting systems.

#### Useful Economic Life:

Purchase of a mechanical harvesting system is usually a capital investment.<sup>4/</sup> A capital asset contributes to a productive process throughout its useful economic life. In the case of a harvesting system, productive life may be shortened by machine use resulting in wear and new technology leading to obsolescence of the system. Useful machine life generates a flow of cost over a number of years and thus a decision model must account for the flow of costs associated with each alternative system.

---

<sup>4/</sup> In some instances, mechanical systems are purchased to satisfy consumption motives or needs of the individual.

## Harvesting System Productivity:

Harvesting system productivity is defined as output of bulk bins of fruit per unit of time.<sup>5/</sup> This measure of productivity is distinguished from measures of economic efficiency.<sup>6/</sup>

Seamount (1969) concludes that among individuals picking from man-positioner harvesters, picking rates differ significantly. He concluded that pickers with the greatest picking variability are influenced more by the method of harvesting, i.e., ladder and bag method versus mechanical aid method, than are those individuals with less variable picking rates. That is, the more skilled, faster pickers tend to be significantly less variable in picking rate and influenced less by method of harvesting. It is assumed that the more skillful and responsible individuals would be assigned to operate expensive man-positioned systems. Thus, this source of system productivity variability is reduced. Furthermore, for those mechanical harvesting systems not utilizing human labor for the picking function, variability caused by a human operator may not be a significant factor.

Climatic conditions, such as rainfall or extreme heat during the harvest season, may contribute significantly to the variability of a harvesting system's productivity. For example, effective movement of mechanical devices would be impaired on rain-softened ground. Performance of human labor and machines can be impeded by climatic conditions. Similarly, extreme topographic features such as hilly, uneven ground may hinder maneuverability, and, therefore, the productivity of a system. These interferences with productivity are generally beyond the control of a producer.

---

<sup>5/</sup> A bulk bin is usually a large wooden container with a capacity of approximately 25 thirty-five pound field boxes of apples.

<sup>6/</sup> Economic efficiency is measured in terms of monetary value of output per unit of monetary input, i.e., the product of time and cost per unit of time. Assuming an apple producer receives the same value for a unit output of fruit and pays the same value per unit of input for any harvesting system he might employ, the economic efficiency of a system will increase if it can produce a unit of output with fewer units of input.

It was assumed that density of fruit in an orchard contributed most to variability of output of the harvesting systems. That is, density of apples on the trees determines the amount of time required to find, move to, reach, pick, and fill a bulk bin. Through availability of apple yield data it was possible to include fruit density in the decision model formulation.

#### The States of Nature

To make a rational choice among two or more apple harvesting systems a producer must have a decision model which considers that apple yields per tree are not uniform from tree to tree within an orchard, and, therefore, neither is the number of trees required to fill each consecutive bin in a harvest operation. Differences in yields among apple trees in an orchard arise from several sources: (1) variety and rootstock combinations, (2) ages of the trees and (3) the nature of individual trees.

#### Yield Differences from Variety-Rootstock Combination:

The apple tree has two major components: (1) the root system and (2) the vegetative system. The root system provides support for the tree and collects necessary nutrients for it to live, grow, and produce. Root systems differ because of their vigors and abilities to perform the life functions of the tree. These inherent differences are characterized by types of rootstocks, that is, some are better suited to different environmental conditions than are others.

Variety refers to the vegetative portion of the apple tree, i.e., Jonathan, Golden Delicious, Red Delicious, or Winesap varieties. The variety grown in a particular region is determined by variety demand of area markets, grower preference, and physical environment conditions.

Mean yields in field boxes per tree in Table 2 are for trees age 9 through 17 years for Golden Delicious and Winesap varieties on seedling rootstock.<sup>7/</sup> Yield data in Table 3 are for trees age 3 through 9 years

---

<sup>7/</sup> A field box is a wooden container, used extensively before the introduction of larger bulk bins, and holds approximately 35 pounds of apples.

for Red Delicious and Golden Delicious varieties on rootstocks including Seedling, Clark, Dwarf, and four types of Malling. Inspection of the rows of yield data in Table 3 reveals the range of yield difference that occurred in an orchard with several variety-rootstock combinations. Note, for example, that the yield for nine-year-old trees varies from a minimum of 3.64 field boxes of apples per tree to a high of 9.61.

Table 2. Mean yields and standard deviations in field boxes per tree by age and variety-rootstock combination a/

Age (years)	Golden Delicious Variety on Seedling rootstock			Winesap variety on Seedling rootstock		
	Mean	Standard deviation	N <sup>b/</sup>	Mean	Standard deviation	N
9	4.15	3.23	197	6.46	4.41	181
10	19.42	7.08	197	14.31	6.44	181
11	11.77	5.20	197	18.53	4.81	181
12	16.35	7.25	197	16.68	6.36	181
13	11.04	5.50	197	14.18	6.00	181
14	19.79	8.47	197	16.89	5.44	181
15	19.10	7.48	197	14.05	8.02	181
16	20.58	8.91	197	21.87	6.07	181
17	20.13	7.22	197	20.49	7.43	181

a/ Means and standard deviations calculated from data obtained from Tree Fruit Research Center, Washington State University, Wenatchee, Washington.

b/ N indicates the number of tree observations used in calculating each mean and standard deviation.

#### Yield Differences from Age of Tree:

Orchard yields vary with respect to age of apple trees. Beginning about age four years, most apple trees produce a harvestable crop. From this age on, a tree generally produces an increasing number of apples each year until a maximum production is reached, after which yields decline.<sup>8/</sup> A yield and

8/ Adverse environmental conditions, whether induced by human or natural forces, may cause crop failure during a specific year.

Table 3. Mean yields and standard deviations in field boxes per tree by age and variety-rootstock combinations a/

Age (years)	Red Delicious variety on Malling I rootstock			Golden Delicious variety on Malling I rootstock			Red Delicious variety on Malling VII rootstock		
	Mean	Standard deviation	N <sup>b/</sup>	Mean	Standard deviation	N	Mean	Standard deviation	N
3	0.11	0.29	110	0.42	0.53	40	0.17	0.36	251
4	0.35	0.34	49	0.00	0.00	7	0.40	0.45	114
5	2.52	2.54	97	2.58	2.61	33	2.00	1.86	221
6	3.16	3.02	110	3.31	3.24	40	2.42	1.85	251
7	5.48	3.39	110	7.22	3.63	34	3.91	2.32	251
8	5.61	5.06	95	4.62	4.20	35	3.60	3.28	215
9	7.17	3.95	110	7.72	3.70	39	4.56	2.77	251

Age (years)	Golden Delicious variety on Malling VII rootstock			Red Delicious variety on Malling XVI rootstock			Golden Delicious variety on Malling XVI rootstock		
	Mean	Standard deviation	N	Mean	Standard deviation	N	Mean	Standard deviation	N
3	0.60	0.57	78	0.04	0.14	74	0.24	0.41	26
4	0.00	0.00	10	0.10	0.15	33	0.00	0.00	2
5	2.48	1.79	67	1.20	1.50	71	2.51	2.65	22
6	2.10	2.20	78	2.40	1.91	80	2.61	2.01	27
7	5.42	2.98	69	4.37	2.85	80	7.32	3.73	24
8	2.61	2.55	66	5.37	4.51	70	4.22	2.90	23
9	5.10	2.97	73	7.47	5.17	80	8.96	4.68	26

(Continued)



Table 3. (Continued)

Age (years)	Red Delicious variety on Malling II rootstock			Golden Delicious variety on Malling II rootstock			Red Delicious variety on Seedling rootstock		
	Mean	Standard deviation	N	Mean	Standard deviation	N	Mean	Standard deviation	N
3	0.03	0.15	105	0.38	0.49	33	0.09	0.25	45
4	0.13	0.16	44	0.00	0.00	6	0.12	0.28	19
5	2.15	2.24	89	2.41	2.41	29	1.77	2.31	40
6	2.98	2.34	104	3.41	2.53	33	2.40	2.20	47
7	5.42	3.10	104	5.66	2.65	26	4.64	3.05	46
8	5.41	4.79	97	5.11	4.73	29	4.30	2.99	39
9	5.87	3.55	104	7.09	3.32	33	5.58	4.13	45

---

Age (years)	Golden Delicious variety on Seedling rootstock			Red Delicious variety on Clark Dwarf rootstock			Golden Delicious variety on Clark Dwarf rootstock		
	Mean	Standard deviation	N	Mean	Standard deviation	N	Mean	Standard deviation	N
3	0.34	0.48	17	0.03	0.11	11	0.79	0.95	19
4	0.00	0.00	3	0.00	0.00	3	0.00	0.00	3
5	3.21	3.78	17	0.90	1.28	9	1.89	2.69	19
6	4.84	2.54	17	2.08	1.71	11	2.04	1.91	19
7	8.42	3.47	14	3.43	2.00	11	4.15	3.49	15
8	7.13	5.03	15	3.77	2.42	9	3.62	4.71	16
9	9.61	4.33	17	3.66	2.02	11	3.64	2.74	18

a/ Means and standard deviations calculated from data obtained from Mid-Columbia Experiment Station, Oregon State University, Hood River, Oregon.

b/ N indicates the number of tree observations used in calculating each mean and standard deviation.

age relationship from data in Tables 2 and 3 is characterized in Figure 2 where age in years is measured along the horizontal axis and yield in mean field boxes per tree along the vertical axis.

To show the characteristics of growth curves for different variety-rootstock combinations, linear regression equations were fitted to each set of variety-rootstock data.<sup>9/</sup> The regression coefficients and other statistics are shown in Table 4.

The coefficient (b) on the age variable (x) is the slope of a growth relationship and provides a measure of the annual rate of increase of yield in field boxes per tree. The differences in the slopes listed in Table 4 represent differences in yield growth vigor for the various variety-rootstock combinations. Therefore, in a harvest operation to relate productivity of a harvesting system to sources of varying yields, the age, as well as the variety-rootstock variable should be considered.

#### Yield Variability Due to Nature of Individual Trees:

Trees, like all living things, react uniquely to variations in environmental stimuli and thus yields may vary between trees of the same age and variety-rootstock combination. This fact is exemplified by the standard deviations of yields in field boxes per tree shown in Tables 2 and 3. For example, for age 16 Golden Delicious variety on Seedling rootstock the standard deviation was 8.91, the highest variance (standard deviation squared) observed in the data. The lowest standard deviation was 0.11 for age three Red Delicious variety on Clark Dwarf rootstock, (excluding those age and variety-rootstock combinations showing zero mean yields).

#### Estimation of Density Variability:

Data presented has substantiated that apple yields per tree are not uniform. For this reason, a computer program was developed to determine the mean

---

<sup>9/</sup> Data were not available to estimate nonlinear regression equations.

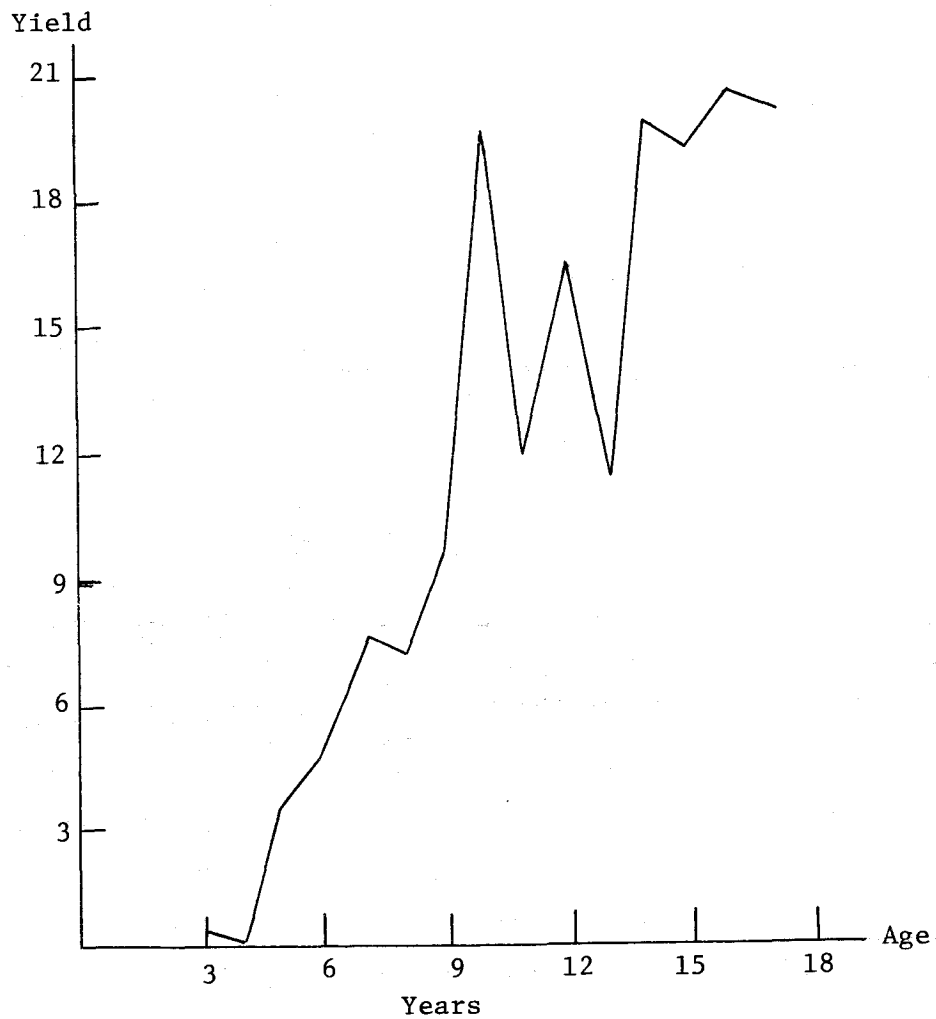


Figure 2. A yield-age growth curve for Golden Delicious variety on seedling rootstock.

Table 4. Linear regression equations and other statistics for yield in field boxes per tree as a function of age in years <sup>a/</sup>

Variety-rootstock combination	Constant (a)	Slope (b)	R-square	F-value
Golden Delicious variety on Seedling rootstock	-4.1359	1.4252	0.1782	383.98
Winesap variety on Seedling rootstock	-0.4071	1.1676	0.1611	312.39
Red Delicious variety on Malling I rootstock	-3.7163	1.2113	0.3616	384.55
Golden Delicious variety on Malling I rootstock	-3.3024	1.1924	0.3488	121.04
Red Delicious variety on Malling VII rootstock	-2.0058	0.7496	0.3226	739.20
Golden Delicious variety on Malling VII rootstock	-1.3971	0.6963	0.2376	136.83
Red Delicious variety on Malling XVI rootstock	-4.6306	1.2753	0.3956	316.84
Golden Delicious variety on Malling XVI rootstock	-4.2136	1.3449	0.4015	99.28
Red Delicious variety on Malling II rootstock	-3.2635	1.0800	0.3480	344.22
Golden Delicious variety on Malling II rootstock	-3.1499	1.1212	0.3826	115.90
Red Delicious variety on Seedling rootstock	-3.0186	0.9641	0.3417	144.84
Golden Delicious variety on Seedling rootstock	-4.4974	1.5791	0.4524	80.97
Red Delicious variety on Clark Dwarf rootstock	-2.1979	0.7083	0.4201	45.64
Golden Delicious variety on Clark Dwarf rootstock	-0.9265	0.5591	0.1358	16.81

<sup>a/</sup> Yield in field boxes per tree =  $f(\text{age of tree in years})$ , i.e.,  $y = atbx$  where  $y$  is the yield in field boxes per tree,  $a$  is the constant,  $b$  is the slope, and  $x$  is the age of the tree in years.

trees per bin and trees per bin variance needed to simulate realistic orchard harvesting yield conditions.<sup>10/</sup> Tables 5 and 6 show means of trees per bin and corresponding standard deviations for the Wenatchee and Hood River data related in Tables 2 and 3. These values were calculated by the trees per bin simulation program.

Each cell in Tables 5 and 6 specified by tree age and variety-rootstock combination was calculated on the basis of 100 bins. The program simulated picking for each yield distribution until 100 bins were filled.<sup>11/</sup> From these tables, information is available to define at least some bounds on the extent of density variability within an orchard. For example, the lowest density was represented by the highest number of trees per bin, i.e., 24.72 trees per bin for the three-year-old Red Delicious variety on Malling VII rootstock.<sup>12/</sup> The highest density was represented by 1.11 trees per bin for age 16 Winesap variety on Seedling rootstock.<sup>13/</sup> However, since three- to four-year-old trees are of questionable harvesting value, a preferred lower bound on density may be 17.09 trees per bin for age five Red Delicious variety on Clark Dwarf rootstock. A realistic estimate of the range of orchard fruit density varied from 1.11 to 17.09 trees per bin. Thus, the important reality of orchard yield variability among

---

<sup>10/</sup> The trees per bin computer simulation program and instructions for its operation are given in Rudkin (1971).

<sup>11/</sup> The distributions of field boxes of apples per tree implied by the simulated results in Tables 5 and 6 were compared with the actual distributions. A Chi-square test showed that there were no significant differences between the simulated and actual distributions.

<sup>12/</sup> For those cells in Tables 5 and 6 specifying an age and variety-rootstock combination with mean density of 25 and zero standard deviation, all field box yields per tree were between 0.0 and 2.0 in each frequency distribution.

<sup>13/</sup> In comparing Tables 5 and 6 with Tables 2 and 3, note the respective values of mean field boxes per tree and mean trees per bin move in opposite directions in all cases. That is, as field box yield per tree rises, the number of trees per bin falls. Thus, the lowest trees per bin measurement represents the highest yielding trees in field boxes; whereas, the highest trees per bin measurement represents the lowest yielding trees in field boxes.

Table 5. Mean trees per bin and standard deviations by age and variety-rootstock combinations <sup>a/</sup>

Age (Years)	Golden Delicious variety on Seedling rootstock		Winesap variety on Seedling rootstock	
	Mean	Standard deviation	Mean	Standard deviation
9	5.70	1.71	3.87	1.21
10	1.27	0.37	1.73	0.53
11	2.09	0.55	1.34	0.29
12	1.47	0.56	1.43	0.40
13	2.13	0.57	1.74	0.52
14	1.31	0.46	1.39	0.33
15	1.45	0.52	1.72	0.67
16	1.13	0.38	1.11	0.24
17	1.23	0.31	1.16	0.34

<sup>a/</sup> Means and standard deviations of trees per bin were calculated by a trees per bin computer simulation program.

Table 6. Mean trees per bin and standard deviations by age and variety-rootstock combinations a/

Age (years)	Red Delicious variety on Malling I rootstock		Golden Delicious variety on Malling I rootstock		Red Delicious variety on Malling VII rootstock	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
3	24.62	0.77	23.72	1.41	24.72	0.68
4	25.00	0.00	25.00	0.00	24.64	0.81
5	9.42	2.63	9.58	2.98	12.10	3.00
6	8.16	2.70	6.57	2.08	9.87	2.54
7	4.79	1.33	3.29	0.98	6.07	1.42
8	4.37	1.69	5.47	2.05	6.42	2.00
9	3.38	0.99	3.14	0.81	5.23	1.37

Age (years)	Golden Delicious variety on Malling VII rootstock		Red Delicious variety on Malling XVI rootstock		Golden Delicious variety on Malling XVI rootstock	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
3	23.84	1.47	25.00	0.00	25.00	0.00
4	25.00	0.00	25.00	0.00	25.00	0.00
5	9.89	2.03	14.59	3.01	9.39	3.13
6	10.26	2.68	9.85	2.15	9.44	1.82
7	4.71	1.17	5.32	1.44	3.38	0.91
8	9.57	2.65	4.41	1.37	5.82	1.65
9	4.71	1.16	3.36	1.32	2.64	0.74

(Continued)

Table 6. (Continued)

Age (years)	Red Delicious variety on Malling II rootstock		Golden Delicious variety on Malling II rootstock		Red Delicious variety on Seedling rootstock	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
3	25.00	0.00	25.00	0.00	25.00	0.00
4	25.00	0.00	25.00	0.00	25.00	0.00
5	10.37	3.12	9.88	2.86	11.76	3.22
6	7.90	2.00	6.97	1.75	9.88	2.50
7	4.88	1.20	4.64	1.12	4.96	1.47
8	4.47	1.69	4.37	1.84	5.84	1.65
9	4.06	1.30	3.61	0.83	4.27	1.48

---

Age (years)	Golden Delicious variety on Seedling rootstock		Red Delicious variety on Clark Dwarf rootstock		Golden Delicious variety on Clark Dwarf rootstock	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
3	25.00	0.00	25.00	0.00	20.32	2.48
4	25.00	0.00	25.00	0.00	25.00	0.00
5	7.56	2.67	17.09	2.32	10.46	2.91
6	4.99	1.09	11.19	2.31	10.36	2.18
7	2.93	0.72	6.59	1.44	5.46	1.53
8	3.69	1.13	6.36	1.42	6.25	2.92
9	2.51	0.65	6.88	1.47	6.56	1.62

a/ Means and standard deviations of trees per bin were calculated by a trees per bin computer simulation program.



apple trees can be identified in a decision model when the states of nature, characterized as fruit density, are continuous within this range.

### The Payoffs

As indicated in the discussion of decision theory framework, the interaction of a particular alternative and occurrence of a specific state of nature will result in a payoff to the decision maker.

### Total Cost of a Harvesting System:

The cells in the payoff table for the harvesting system decision specified by action-state of nature combinations contain present values of total cost. Total cost is comprised of fixed and variable costs. For this study, variable cost is defined as a function of the amount of time a system is operated. To calculate variable cost for the payoff component of the decision model, it is first necessary to know the amount of time a system will operate to pick an orchard. The time needed, and, therefore, total variable cost depend upon: (1) the number of fruit-yielding trees in an orchard, (2) fruit density in trees per bin, and (3) productivity of a system in terms of time per bin. Therefore, if productivities of alternative systems are inherently different, and if orchard yields vary, variable cost from system to system will differ for the same orchard. In addition, if yields of a group of trees vary from one year to the next, and if productivities of alternative systems depend upon yield conditions, the respective variable cost of each system will vary from one year to the next over the useful life.

Productivity of a harvesting system is measured in time per bin, therefore, variable cost can be expressed in terms of cost per unit of time. Total variable cost for a harvesting system is the product of variable cost per unit of time and total time required to pick an orchard. Variable cost per unit of time for a system for a specific time period may be expressed as:

$$VCT = (OPVC + LVC) \text{ when}$$

OPVC = operating cost per unit time, fuel, oil, etc.

LVC = labor cost per unit time.<sup>14/</sup>

---

<sup>14/</sup> Equation symbols throughout this report correspond to the nomenclature used in the computer programs of this study.

Harvesting systems generally may have a useful life extending beyond one year, therefore a stream of variable cost will be incurred by their operation throughout their useful life. Total variable cost for the life of a system is given by the expression:

$$TVC = TVC_1 + TVC_2 + \dots + TVC_T \quad \text{when}$$

$$TVC_i = (TIME_i) (VCT_i) \quad \text{and when}$$

$TIME_i$  = total time to pick an orchard for each year of the useful life of a system,  $i=1, \dots, T$ ;

$VCT_i$  = variable cost per unit of time for each year of the useful life of a system,  $i=1, \dots, T$ .

The variable,  $TIME$  for a given year of the useful life of a system, depends upon the size of the orchard(s), productivity of the system, and yields of trees in the orchard(s), and may be expressed:

$$TIME_i = TBIN_1 + TBIN_2 + \dots + TBIN_N \quad \text{when}$$

$TBIN_j$  = the time it takes to harvest each bin for the  $j=1, \dots, N$  bins in an orchard.

Time per bin ( $TBIN$ ) comes from a relationship between a system's productivity measured in time per bin, and fruit density of an orchard measured in trees per bin. Figure 3 represents a conjectured functional relationship between a system's productivity measured in time per bin along the vertical axis and trees per bin along the horizontal axis. The relationship in Figure 3 indicates that as the number of trees per bin increases (decreasing fruit density), the time required to harvest a bin of fruit increases. Given the number of trees picked to fill each bin during a harvest operation, the time to fill each of these bins can be determined from an estimated relationship similar to that in Figure 3. Thus, the total time to pick an orchard in a given year ( $TIME$ ) is the summation of all the time per bin observations taken from the functional relationship.

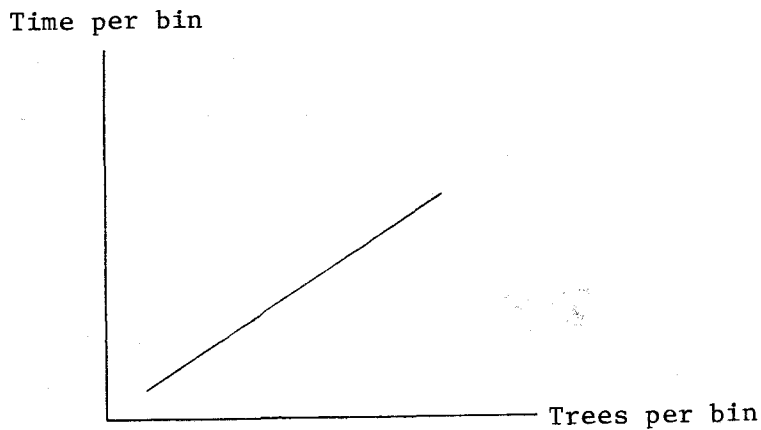


Figure 3. A conjectured functional relationship between time per bin and trees per bin.

Total cost (TC) of a harvesting system over its useful life is the summation of total variable cost (TVC) and fixed cost of the system (FC). That is:

$$TC = FC + TVC \text{ when}$$

TVC = total variable cost;

FC = fixed cost which is composed of the initial price of a system (INV) less the downpayment on its purchase (DNPAY) plus the finance cost of installment buying (COSTF).

Present Value of Total Cost:

Total variable cost (TVC) is the summation of variable cost for each year of useful life ( $TVC_1$ ). Fixed cost (FC) is distributed over an amortization period (M) of installment purchase. Since these costs occur in different time periods throughout the useful life of the system, streams of costs for different harvesting systems may be put on a comparable basis by calculating present values. Present value of total cost for a harvesting system over its entire useful life may be expressed:

$$\begin{aligned}
 PVTC = & DNPAY + TVC_1 + \frac{YFIXC}{1+r} + \\
 & \frac{TVC_2}{1+i} + \frac{YFIXC}{(1+r)^2} + \dots + \\
 & \frac{TVC_M}{(1+i)^{M-1}} + \frac{YFIXC}{(1+r)^M} + \dots + \\
 & \frac{TVC_T}{(1+i)^{T-1}} \quad \text{when}
 \end{aligned}$$

yearly fixed cost is  $YFIXC = \frac{INV - DNPAY + COSTF}{M}$  ;

M = amortization period of an installment purchase;

T = expected useful life of a system,  $i=1, \dots, M \dots, T$ ; and

i and r are discount rates.

Present value of total cost for the useful life of a harvesting system is equal to the downpayment on its purchase, plus the discounted variable cost for each year of system use at discount rate i, plus the discounted annual fixed cost over an amortization period at discount rate r.<sup>15/</sup>

The Decision Criterion:

Given present values of total cost for two or more alternative harvesting systems, the decision criterion for selecting one system could be minimization of expected present value of total cost. However, when the decision criterion is minimization of disutility, then expected values and variances of present value of total cost must be found. In the last section of this report a computer program which simulates the operation of each alternative system over its useful life (for an orchard with the distribution of states of nature continuous and age and variety-rootstock combination vary) is described. This program calculates the distribution of present value of total cost for the useful life of each alternative harvesting system from which expected values and variances can be determined. But first, the relationship between time per bin and trees per bin needs to be given further elucidation.

#### ESTIMATION OF THE RELATIONSHIP BETWEEN SYSTEM PRODUCTIVITY AND FRUIT DENSITY

From the previous discussion it is apparent that if a productivity-density relationship is known for each of several alternative systems, present value of total cost can be calculated for each action-state of nature combination. In the following sections a sample size formula and procedure are presented which make estimation possible from performances test data of the productivity density relationships. A discussion of technical efficiency and economic efficiency will prepare the groundwork for the formula and procedures for estimating these relationships.

---

<sup>15/</sup> The discount rate r represents the rate at which funds can be borrowed; the discount rate i represents the opportunity rate at which funds can be invested.

## Technical Efficiency and Economy Efficiency

Technical efficiency is the measure of harvesting system productivity and is defined as the time required to fill a bulk bin with apples. This measure of system productivity embodies no accounting for monetary value of the input (time) or the output (bins of fruit). Economic efficiency is the appropriate measure when monetary values are attached to inputs and outputs. To illustrate the distinction between technical efficiency and economic efficiency, an example will be presented to clarify their roles in the decision problem.

Suppose the labeled  $S_L$  in Figure 4 describes the functional relationship between the time per bin and the trees per bin measure of fruit density for an orchard harvested by a conventional ladder harvesting system  $S_L$ . Furthermore, suppose this conventional ladder productivity-density functional relationship is a standard against which all other alternative harvesting methods are to be compared. The line labeled  $S_A$  in Figure 4 depicts the productivity-density relationship for an alternative harvesting system  $S_A$ .

Figure 4 indicates that at a fruit density of  $D$ , the conventional ladder method  $S_L$  required  $TPB_L$  time per bin to fill one bin; at the same fruit density the alternative harvesting method  $S_A$  required  $TPB_A$  time per bin. ( $TPB_A < TPB_L$ ). It can be concluded that  $S_A$  is the better harvesting method because it has a

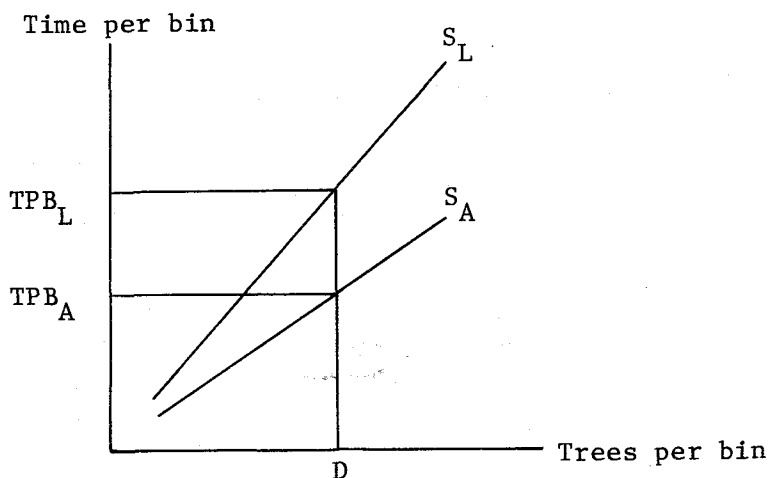


Figure 4. Hypothetical functional relationships between time per bin and trees per bin for systems  $S_L$  and  $S_A$ .

higher technical efficiency at all levels of fruit density.<sup>16/</sup> The higher technical efficiency of  $S_A$  does not necessarily make it the most economically efficient system, i.e., technical efficiency does not reflect relevant costs or returns.

To illustrate, assume an apple producer receives the same value for a unit of output for any harvesting system he employs, the economic efficiency of  $S_A$  can be less than that of  $S_L$  if the cost per unit of input for  $S_A$  more than offsets its greater technical efficiency.<sup>17/</sup> However, for the relationship depicted in Figure 4, if the cost per unit of time is the same for both  $S_L$  and  $S_A$ , the alternative system's economic efficiency can be greater than that of the ladder system, and thereby agree with the conclusion reached by comparing only technical efficiencies. However, technical efficiency can be made a preliminary basis for identifying different systems. That is, given the productivity-density relationship for  $S_L$  as a comparative standard, a relationship for  $S_A$  can be generated such that the economic efficiency of each system is the same. For example, suppose the conventional ladder system has the linear productivity-density relationship described in Figure 5 by the line labeled  $S_L$ . Furthermore, suppose variable cost per unit time (minutes) is a labor cost of \$0.05 per minute. In addition, assume one bin of fruit is picked by  $S_L$  requiring 60 minutes to harvest from a group of trees with fruit density of two trees per bin. A simple measure of technical efficiency is 60

---

<sup>16/</sup> At the origin, fruit density becomes infinitely great, i.e., an infinitesimally small portion of a tree yields one bulk bin. Also at this point of infinite fruit density, the time required to fill one bin is infinitesimally small. Of course, in reality this situation will never occur, thus the definitional problem of zero fruit density and zero time per bin has no practical impact on the discussions to follow, and does not invalidate the relationships depicted in Figure 4.

<sup>17/</sup> It is assumed throughout the analysis of this study that two harvesting systems will provide picked fruit of equal quality and, hence, the price of the fruit does not enter into the calculations. However, if there were a difference in quality of fruit due to the harvesting system, then the reduction in price due to a loss in quality could be an additional cost to be accounted for in the cost calculations presented here.

To date there have not been sufficient tests of the influence of the many alternative experimental harvesting methods on fruit quality.

minutes per bin; \$3 per bin is a simple measure of economic efficiency. The technical efficiency of the alternative harvesting system  $S_A$  is unknown before testing. For  $S_A$  to have the same economic efficiency as  $S_L$ ,  $S_A$  must exhibit by performance testing: (1) the same technical efficiency and variable cost per unit of time as  $S_L$  or (2) a higher technical efficiency ( $TPB_A < TPB_L$ ) with compensating higher variable cost per unit of time, or (3) a lower technical efficiency ( $TPB_A > TPB_L$ ) with compensating lower variable cost per unit of time.

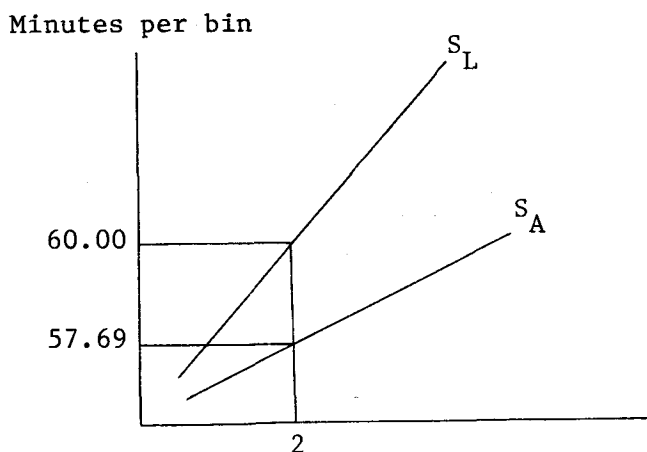


Figure 5. Technical efficiencies of harvest systems  $S_L$  and  $S_A$  at equivalent economic efficiencies.

In reference to the example, suppose system  $S_A$  has the same labor cost per unit of time as  $S_L$ , i.e., \$0.05 per minute, but an additional cost for operating per minute of \$0.002. Variable cost per unit of time for  $S_A$  is \$0.052. For  $S_A$  to have the same economic efficiency as  $S_L$  (\$3 per bin), the technical efficiency  $S_A$  must exhibit can be determined by the following expression:

$$TPB_A = \frac{(TPB_L)(VCT_L)}{VCT_A} \quad \text{when}$$

$TPB_A$  = time (minutes) per bin, alternative harvesting system  $S_A$ ,

$TPB_L$  = time (minutes) per bin, conventional labor system  $S_L$ ,

$VCT_L$  = variable cost per unit of time (minutes), conventional ladder system  $S_L$ , and

$VCT_A$  = variable cost per unit of time (minutes), alternative harvesting system  $S_A$ .

The unknown in this expression is  $TPB_A$ , time per bin for the alternative harvesting system  $S_A$  at a fruit density of two trees per bin. Solution of the equation gives the time per bin that will equate the economic efficiency of  $S_A$  to that of  $S_L$  in terms of variable cost:

$$TPB_A = \frac{(60.0) (\$.05)}{\$.052} = 57.69 \text{ minutes per bin.}$$

That is, system  $S_A$  must pick one bin of fruit in 57.69 minutes per bin at fruit density of two trees per bin and variable cost per minute of \$.052 to have the same economic efficiency of  $S_L$  in terms of variable cost. In short, both systems now have the same economic efficiency of \$3 per bin, but  $S_A$  must pick at a rate of at least 2.31 (60.0 - 57.69) minutes per bin faster than  $S_L$  from trees exhibiting a fruit density of two trees per bin to be an economically viable alternative action to  $S_L$ .

Variable cost per unit of time determines the minimum magnitude of technical efficiency difference that must exist between two systems to equate their economic efficiencies. It is a minimum magnitude because fixed cost has not been considered. That is, the magnitude of a difference between technical efficiencies will have to be larger than a minimum difference determined only with respect to variable cost. However, fixed cost is accounted for in the calculation of payoff for a particular action-state of nature combination. Later fixed cost will be included in the calculation of present value of total cost for the useful life of each alternative action in the decision problem.

#### Difference Between the Slopes

To estimate productivity-density relationships of apple harvesting systems, it is necessary: (1) to determine how large a difference between two relationships is important, and (2) to determine the number of observations needed in a performance test of the system to provide statistical confidence that a specified difference can be detected.



In the previous example it was shown that  $S_A$  must pick at a rate of no less than 2.31 minutes per bin faster than  $S_L$  to maintain an equivalent economic efficiency in terms of variable cost. Figure 5 expresses this difference in terms of the relative slopes of systems  $S_L$  and  $S_A$ . The slope of line  $S_L$  is 30 minutes (per bin) per tree (per bin). Given the difference between the technical efficiencies of  $S_L$  and  $S_A$  is 2.31 minutes per bin at a fruit density of two trees per bin, and variable cost per unit of time is constant,  $S_A$  must have a slope of 28.84 minutes (per bin) per tree (per bin) to be an equivalent system to  $S_L$  in terms of variable cost. The difference between their respective slopes is 1.16 minutes (per bin) per tree (per bin). For purposes here, the specified difference to equate the two systems on an economic basis is called  $d_s$ . Thus the size of  $d_s$  is a difference against which any estimate of the actual difference between two systems can be tested.

The actual or true difference is  $d_o = B_L - B_A$  when  $B_L$  is the actual or true slope of the productivity-density line for system  $S_L$ , and  $B_A$  is the actual or true slope for system  $S_A$ . The actual difference  $d_o$  is not known and must be estimated by performance tests of the harvesting systems. The estimated difference  $d_o$  can be statistically tested against  $d_s$  to determine if  $d_o$  is equal to or greater than  $d_s$ . If  $d_o$  is equal to  $d_s$ , the two systems are equivalent provided their respective fixed costs are similar. However, when comparing a conventional ladder system to a mechanical harvesting system, the fixed cost of the ladder system will be less than that of the mechanical system, and hence  $d_o$  must be greater than  $d_s$  if the mechanical system is to be an economically viable action. Thus, the null hypothesis in experimental performance tests of a conventional harvesting system and a mechanical system is  $d_o \leq d_s$ . Should the null hypothesis be rejected, the alternative hypothesis that  $d_o > d_s$  is accepted and the conclusion follows that the mechanical harvesting system is a possible harvest alternative.

#### The Sample Size Formula for Estimating Productivity-Density Relationships and for Detecting a Significant Difference

The questions in performance testing of harvesting systems is: how many observations of time per bin and trees per bin are needed to be statistically confident that a significant difference between  $d_o$  and  $d_s$  exists? A formula

for calculating performance test sample size in numbers of bins for each harvesting system has been developed.<sup>18/</sup> The formula assumes the null hypothesis  $d_o \leq d_s$  is not true. Thus the purpose of the formula is to provide a large enough sample size to statistically detect the alternative hypothesis  $d_o > d_s$ .

The formula is:

$$N = \frac{2 \sigma^2}{SS} \left( \frac{P^2}{d_o - d_s} \right) \quad \text{when}$$

$N$  = sample size in numbers of bins for each harvesting system;

$\sigma^2$  = variance about a linear regression line, time per bin =  $f$  (trees per bin). In application an estimate of  $\sigma^2$  is used;

$SS$  = average of the squares of the independent variable, trees per bin. In application an estimate of  $SS$  is used;

$P = Z_\alpha + Z_\beta$  when  $Z_\alpha$  is the normal deviate corresponding to the significant level  $\alpha$ , and  $Z_\beta$  is the normal deviate corresponding to the power of the test;

$d_o$  = actual or true difference between the slopes of two linear productivity-density functions;

$d_s$  = specified difference between the slopes of two linear productivity-density functions as calculated by equating the economic efficiencies in terms of variable cost.

#### Example of Sample Size Determination

To demonstrate calculation of a sample size, data adapted from previous performance tests and data from Table 7 will be used (Berlage, et al., 1966). The value of  $\hat{\sigma}^2$ , an estimate of  $\sigma^2$ , was found to be 103.32 when a linear

<sup>18/</sup> The formula was developed by Kenneth Burnham of the Department of Statistics, Oregon State University. See Appendix for the theoretical derivation of the formula.

regression line was fitted to pooled data from the four commercial man-positioner aids investigated by the Berlage (1966) study.<sup>19/</sup>

The value of SS was 22.11, that is,

$$SS = \frac{7.56^2 + 4.99^2 + 2.93^2 + 3.69^2 + 2.51^2}{5} = 22.11 .$$

The values in the numerator correspond to the means of trees per bin for Golden Delicious variety on Seedling rootstock trees five through nine years old as can be found in Table 7. The value in the denominator is the number of terms in the numerator and provides the average of the squares. Values of SS are specific to the individual apple orchard on which harvesting systems are to be tested, and are estimated for the specific composition of age and variety-rootstock combinations present. Two hypothetical harvest systems are performance tested using the above input data.

For this example,  $P = 1.645 + 1.280 = 2.925$ . The value of  $Z_\alpha$  is 1.645 and corresponds to the upper 95 percent of the normal distribution with mean zero and variance equal to one.<sup>20/</sup> The value of  $Z_\beta$  is 1.280 and corresponds to the upper 90 percent of the normal distribution with mean zero and variance equal to one.<sup>21/</sup>

The value  $d_s$  was derived by calculating the slopes of the linear productivity-density functions of  $S_L$  and  $S_A$  by the following formulas:

---

<sup>19/</sup> These mechanical harvesting systems were not performance tested over a range of fruit density, but this study provided the only data from which an approximation of  $\sigma^2$  could be obtained. Future performance tests following the guidelines of this report should provide better estimates of  $\sigma^2$ .

<sup>20/</sup> The significant level  $\alpha$  is the probability that the null hypothesis  $d_o \leq d_s$  will be rejected when it is true. Tables values are given in Snedecor and Cochran (1967) p. 548.

<sup>21/</sup> The power of the test  $(1-\beta)$  corresponds to  $Z_\beta$  and is the probability of rejecting the null hypothesis  $d_o \leq d_s$  when the alternative hypothesis  $d_o > d_s$  is true. Tables values are given in Snedecore and Cochran (1967) p. 548. A table of  $Z_\alpha + Z_\beta$  values for different powers and one-tailed significance levels is given in Snedecor and Cochran (1967) p. 113.

$$B_L = \frac{TPB_{L1} - TPB_{L2}}{D_1 - D_2} ;$$

$$B_A = \frac{TPB_{A1} - TPB_{A2}}{D_1 - D_2} ;$$

and computing the difference  $B_L - B_A = d_s$ .

The values of  $TPB_{L1}$  and  $TPB_{L2}$  used in this example were taken from the linear regression of time per bin on trees per bin fitted to the data adapted from the Berlage (1966) study. The values of  $TPB_{A1}$  and  $TPB_{A2}$  were calculated by the formula

$$TPB_A = \frac{(TPB_L)(VCT_L)}{VCT_A}$$

as presented earlier. The values of  $D_1$  and  $D_2$  are two density levels corresponding to respective time per bin ( $TPB_{L1}$  and  $TPB_{L2}$ ) measurements. The slope  $B_L$  of system  $S_L$  was 67.29 minutes (per bin) per tree (per bin). Assuming variable cost per unit of time  $VCT_L$  for the ladder system  $S_L$  equals \$0.0333 per minute and variable cost per unit of time  $VCT_A$  for the alternative system  $S_A$  equals \$0.0375 per minute, the time per bin  $TPB_A$  for system  $S_A$  must be at least 59.81 minutes (per bin) per tree (per bin).<sup>22/</sup> Therefore,  $d_s$  is equal to 7.48, (67.29 - 59.81).

The value of  $d_o$  is the true difference between the slopes of the  $S_L$  and  $S_A$  functions. The true difference will have to be larger than the specified  $d_s$  by an amount which would be expected to cover the average fixed cost per minute of operation of the alternative system  $S_A$ . The average fixed cost per minute is equal to the total fixed cost divided by the total time (minutes) that the alternative system is expected to operate throughout its useful life. For this example, it was assumed that average fixed cost was \$0.0037 per minute for

---

<sup>22/</sup> Labor cost for system  $S_L$  was assumed \$2 per hour. Labor cost for system  $S_A$  was also assumed to be \$2 per hour, but an additional operating cost of \$0.0042 per minute or \$0.25 per hour was included. Values for variable cost per unit of time were obtained from Burkner, et. al. (1968-69).

system  $S_A$ .<sup>23/</sup> Including the fixed cost in the denominator of the formula for  $TPB_A$  implies that the slope of  $S_A$  must be 54.44 minutes (per bin) per tree (per bin). Therefore,  $d_o = 67.29 - 54.44 = 12.85$ , that is,  $d_o$  must exceed  $d_s$  by 5.37 minutes (per bin) per tree (per bin) to be a true economic alternative action.

The sample size for this example is calculated using the sample size formula presented above:

$$N = \frac{(2) (103.32)}{22.11} \left( \frac{2.925}{12.85 - 7.48} \right)^2 = 2.77 \text{ or}$$

3 bins for each system to be tested.

### Procedures for Sampling and Testing Hypotheses

The purpose of performance testing is to estimate the productivity-density relationships for apple harvesting systems. The sample size formula provides a means for determining the number of bins to harvest to insure that the experimenter can be statistically confident that he can detect the alternative hypothesis that  $d_o > d_s$ .

---

<sup>23/</sup> Fixed cost is comprised of the initial price of a system less the down payment on purchase, plus the cost of financing an installment purchase. For performance testing, the research will need to make an approximation of the expected number of bins a specific system will be capable of harvesting in its useful life and an estimate of time per bin. The number of bins expected to be harvested depends on the number of trees in an orchard, yields of the trees, and useful life of the system. For the example above, it was assumed the price of the alternative system  $S_A$  is \$3,000, \$1,000 of which is attributed to the harvest operation. The remaining \$2,000 is allocated to pruning and thinning operations. Furthermore, it was assumed that the down payment on purchase is \$200, and the cost of financing an installment purchase on \$800 is \$160, assuming a ten percent interest rate for a two-year amortization period. The present value of fixed cost is equal to:

$$\$200 + \frac{\$480}{1.10} + \frac{\$480}{(1.10)^2} = \$1033.06.$$

It was assumed system  $S_A$  will harvest 4,738 bins of apples in its five year useful life. Therefore, the average fixed cost per bin is  $\frac{\$1033.06}{4,738}$

equals \$0.21804 per bin. It was further assumed for this example that  $S_A$  can harvest at a rate of one bin per 58.81 minutes, and, hence, average fixed cost per minute is  $\frac{\$0.21804}{58.81}$  equals \$0.0037.

Ultimately, use of a sample size calculated by the formula for performance data collection will depend on the size of an orchard test site, expected yield of the test site orchard, and the cost of taking the sample. Extreme values for the four variables in the formula may cause sample size to become extremely large, thus, a sample size would have to be one that would not only satisfy statistical requirements but also one that could be satisfied adequately by an orchard's size. The cost of data collection and analysis may place an upper bound on the sample size, quite possibly a sample size smaller than that indicated by the formula. Thus, there is usually a trade-off between statistical confidence and the need to stay within the limits imposed by budgets, time, and effort. In practice, the experimenter must decide whether the value of the information to be obtained from collecting a "statistically sound" sample is sufficient to warrant the cost of collecting it.

#### Allocation of a Sample Size and Collection of Observations

The physical conduct of performance tests of harvesting systems must be preceded by detailed planning of every aspect of the data collection technique. Procedures to secure uniformity in conduct of performance tests from system to system must be defined. Control must be exercised over external influences which may distort measurement. Observations and systematic measuring techniques must be developed to prevent gross errors in the collection of data.

The productivity-density relationship of an apple harvesting system should show the technical efficiency of the system over a range of fruit density indicative of that which may be encountered by the system throughout its useful life as the apple orchard ages. To obtain usable observations, the sample size must be allocated throughout the test orchard in a manner enabling the collection of time per bin observations over the widest range of fruit density. The authors suggest no general scheme for the allocation of sample size in all orchards. Each orchard is unique in its concentration and range of fruit density. It may be necessary for the experimenter to determine for himself, or rely on others with the ability to judge yields of trees, where in a particular orchard a sample size of bins should be allocated.

The collection of data for the estimation of the productivity-density relationship for a harvesting system will necessitate careful supervision. This means accurate measurement must be taken of the time required to fill each consecutive bin of the sample, and precision provided in the estimation of fractions of trees per bin. Furthermore, extreme care must be taken not to interfere with the normal operation of a system, if accurate measurement of productivity is to be obtained.

After collecting time per bin and trees per bin observations for each harvesting system, a linear regression that describes time per bin as a function of trees per bin can be fitted to the data. It is suggested that regression equations be fitted without intercepts in the preliminary data analysis to facilitate the hypothesis test that the observed difference,  $d_o$ , is significantly less than the specified difference,  $d_s$ , used in the sample size formula. That is, the simplest model should be considered first before further refinements are added.

#### Procedures for Testing Hypotheses

In testing the null hypothesis that the observed difference  $d_o \leq d_s$  versus the alternative hypothesis that  $d_o > d_s$ , the following test statistic is used:

$$t = \frac{\hat{d}_o - d_s}{\sqrt{\frac{2\sigma^2}{SS}}}$$

$\hat{d}_o = \hat{B}_L - \hat{B}_A$  = observed difference between slopes of the linear regression equations of two harvesting systems,

$d_s$  = specified difference between slopes of the linear regression equations of two harvesting systems,

$\hat{\sigma}$  = pooled estimate of variance about the fitted regression equations,

SS = average of the squares of the independent variable, trees per bin from the sampled bins.

If the value of  $t$  exceeds the critical value  $t_{\alpha}$  with  $(n_L + n_A - 2)$  degrees of freedom, the null hypothesis  $d_o \leq d_s$  is rejected at the specified

one-tailed significance level  $\alpha$ .<sup>24/</sup> In the event there is a statistically significant difference, the alternative hypothesis  $d_o > d_g$  is accepted.

Following the test for slope differences two types of "goodness of fit" tests on each linear regression equation should be made: (1) to test whether the productivity-density relationship for a system has a nonzero intercept and (2) to test whether the relationship is nonlinear.

To test the null hypothesis that a linear regression goes through the origin versus the alternative hypothesis that it has an intercept, standard test procedures are available (Snedecor and Cochran (1967) pages 166-167).

To test whether a productivity-density relation is nonlinear, an F-ratio of the mean square due to the reduction in sum of squares from curvilinear regression over the mean square due to deviations from the curvilinear regression is calculated. A significant F-value indicates that the hypothesis of linear regression be rejected, and that there is a significant curvilinearity in the regression (Snedecor and Cochran, 1967, page 455).

#### CALCULATIONS OF PAYOFFS IN THE DECISION MODEL AND SELECTION OF AN ACTION

The computer program to be described in this section simulates the picking of an apple orchard by a harvesting system to obtain estimates of picking time. The computer program calculates expected value and variance of present value of total cost for the useful life of each alternative system to which a decision criterion can be applied.

#### Simulation of a Distribution of Present Value Of Total Cost

Figure 6 presents a flow diagram of the computer program for simulating a distribution of the present value of total cost for an apple harvesting

---

<sup>24/</sup> Since  $\sigma^2$  is estimated, the Student's t-distribution is used. The critical value  $t_{\alpha}$  corresponds to the upper  $(1-\alpha)$  percent point of the Student's t-distribution. The values of  $n_L$  and  $n_A$  corresponds to the number of observations used in the computation of  $B_L$  and  $B_A$ , respectively.



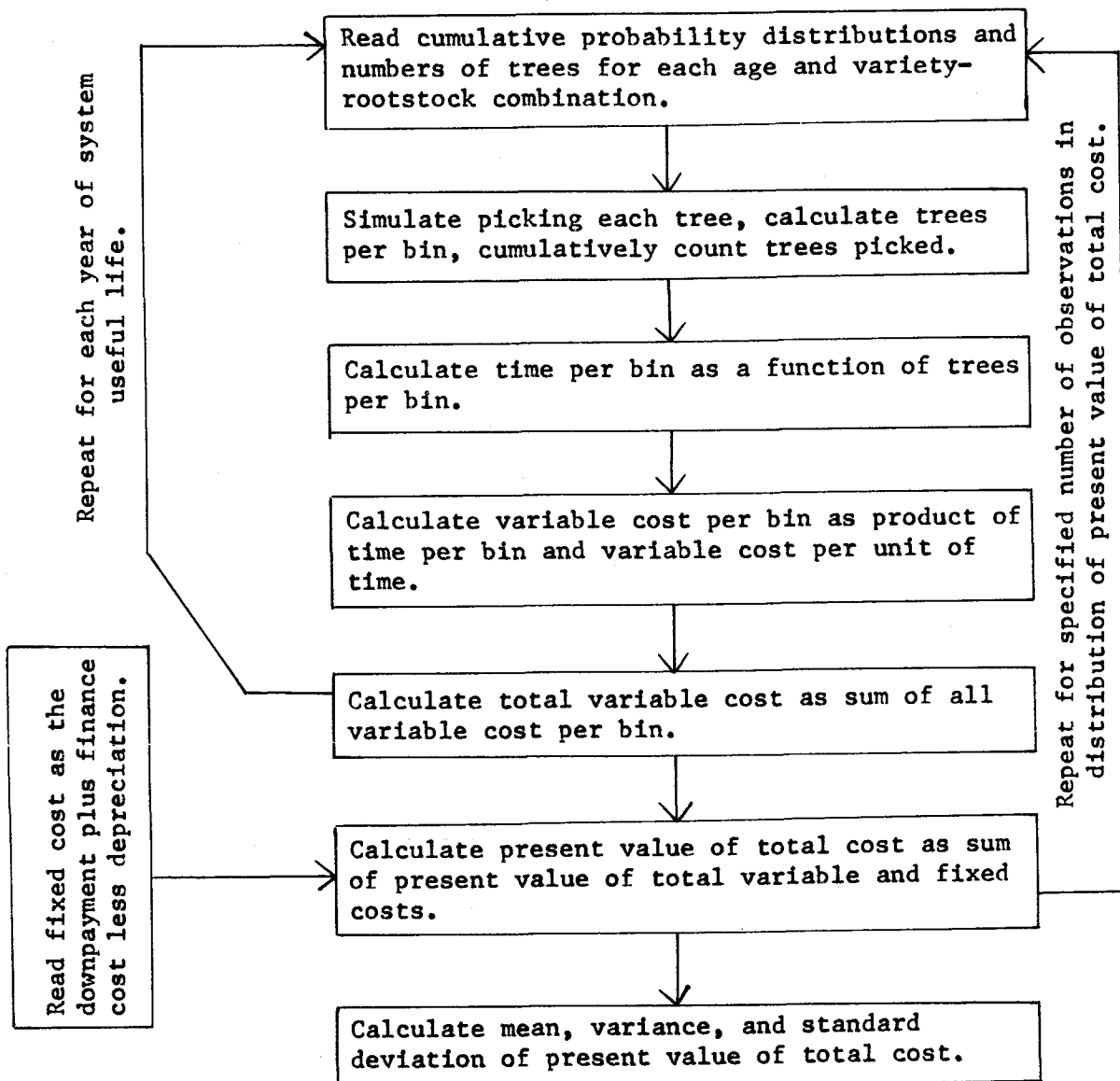


Figure 6. Flow diagram of the computer program for simulating present value of total cost.

system.<sup>25/</sup> The simulation begins by reading a series of data characterizing the particular age and variety-rootstock composition of a specific orchard. These data are similar to those used in the computer program discussed earlier to calculate mean and variance of trees per bin, i.e., cumulative probability distributions of yields in field boxes per tree. The program reads a cumulative distribution for each variety-rootstock combination in an orchard for each year of a harvesting system's useful life. Furthermore, the program will accept any number of age and variety-rootstock combinations, and in addition, it will simulate picking of any specified number of trees for each age and variety-rootstock combination characterizing an orchard.

For example, suppose an orchard in year one of a harvesting system's useful life is composed of 50 three-year-old trees of Red Delicious variety on Clark Dwarf rootstock, 30 trees five years old of Red Delicious variety on Clark Dwarf rootstock, and 50 trees five years old of Golden Delicious variety on Seedling rootstock. Thus, the orchard contains 130 trees of two variety-rootstock combinations and two age groups. The computer program simulates picking, in a random order, each tree of each age and variety-rootstock combination until all 130 trees have been harvested.

The computer program simulates the encounter of apple trees with different yields and the placement of each tree's yield in a 25 field box capacity bulk bin. It calculates the number of trees and fractions of trees which are needed to fill each consecutive bin until a specified total number of trees in an orchard are harvested. With each filled bin, the program calculates from a specified productivity-density relationship the time required to fill each bin i.e.,  $\text{time per bin} = f(\text{trees per bin})$ .<sup>26/</sup> Next, the simulator calculates variable cost for each bin as the product of time required to fill each bin and variable cost per unit of time. This may be expressed as:

---

<sup>25/</sup> The computer simulation program for calculating the distribution of present value of total cost is given in T. H. Rudkin, *op. cit.* Appendix D.

<sup>26/</sup> The computer program contains a random normal deviate generator for calculation of an error term which may be added to or subtracted from a calculated time per bin. A nonzero standard deviation about regression must be read as input to the program to activate calculation of error terms. Furthermore, an initial random number must be read as input to initialize the random number generator which draws normal deviates.

VCB = (TPB) (VCT) when

VCB = variable cost per bin,

TPB = time required to fill a bin,

VCT = variable cost per unit of time for the specified harvesting system.<sup>27/</sup>

After harvesting each of the specified number of trees in an orchard and calculating variable cost for each bin of fruit harvested from these trees, the program calculates total variable cost for the completed harvest operation. That is, the computer program sums all variable cost per bin for all bins harvested from trees in the orchard. Total variable cost for the completed harvest operation in the first year of a system's useful life is expressed as:

$$TVC_1 = \sum_{j=1}^N VCB_j$$

when the orchard yields  $j=1, \dots, N$  bins of fruit. Total variable cost is the summation across variable cost per bin for all age and variety-rootstock combinations specified in an orchard. Thus, total variable cost for the first year of a harvesting system's useful life is estimated. Should a harvesting system's useful life exceed that of one year, the computer program will calculate total variable cost for each year of useful life, i.e., the program will increment the age of each variety-rootstock combination in an orchard by one year and begin again to simulate harvest of the orchard, and calculate total variable cost for the second year. This iterative process will continue until the orchard has been, in effect, harvested once for each year of useful life of the harvesting system.

After simulating the harvest of an orchard for each year and calculating total variable cost for each year, the computer program calculates present value of total cost for a harvesting system for its entire useful life. Present value of total cost for a system over its entire useful life may be expressed:

---

<sup>27/</sup> VCT = (OPVC + LVC) when  
OPVC = operating cost per unit of time for the specified harvesting system,  
LVC = labor cost per unit of time.

$$\begin{aligned}
PVTC = & DNPAY + TVC_1 + \frac{YFIXC}{1+r} + \frac{TVC_2}{1+i} + \frac{YFIXC}{(1+r)^2} \\
& + \dots + \frac{TVC_M}{(1+i)^{M-1}} + \frac{YFIXC}{(1+r)^M} + \dots + \frac{TVC_T}{(1+i)^{T-1}}
\end{aligned}$$

when

PVTC = present value of total cost;

DNPAY = portion of the downpayment on purchase of a system allocated to the harvest operation;

$TVC_i$  = total variable cost for year  $i=1, \dots, M, \dots, T$ ;

T = expected useful life of a system;

i = opportunity rate at which funds can be invested;

r = rate at which funds can be borrowed.

$$YFIXC = \frac{INV - DNPAY + COSTF}{M} \quad \text{when}$$

INV = portion of the initial price of a system allocated to the harvest operation,

COSTF = portion of the finance cost of an installment purchase of a system allocated to the harvest operation,

M = amortization period of an installment purchase.

At this point the computer program has calculated one present value of total cost observation for the useful life of a harvesting system. The objective of the simulation is to generate a distribution of present value of total cost for a system for its useful life on a given orchard. Therefore, the program recycles the entire process to generate a second observation of present value of total cost. That is, it returns to the first year of a system's useful life and harvests the same orchard composition as in the first observation, but in a different random order than that of the first observation, and proceeds to make the calculation for a second observation of present value of total cost. This recycling process continues until a

predetermined number of present value of total cost observations has been generated. Given a distribution, the program calculates a mean, variance, and standard deviation of present value of total cost for a particular system for its useful life in a specified orchard. The following section presents some results of sensitivity tests conducted with this simulation program for hypothetical situations.

### Results of Sensitivity Tests

The purposes of the sensitivity tests were:

- 1) To ascertain the effects of assuming a linear productivity-density relationship for a conventional ladder system  $S_L$  versus an alternative harvesting system  $S_A$ ;
- 2) To demonstrate the effect on variance of present value of total cost of including a standard deviation about regression;
- 3) To ascertain the effects of assuming a curvilinear productivity-density relationship for each of the two systems,  $S_L$  and  $S_A$ ;
- 4) To derive information to demonstrate applications of decision criteria.

A summary of results of sensitivity tests is given in Table 8. Rows 1 and 2 show the results of assuming a linear productivity-density relationship for an alternative harvesting system  $S_A$  and a conventional ladder system  $S_L$ , respectively. The values of the slopes (b) for  $S_A$  and  $S_L$  in rows 1 and 2, respectively, are the same as those used in calculating the example sample size. The variable and fixed costs for the respective systems used in the simulations for the results in Table 8 were also those used in the example.

A value for standard deviation about linear regression was not specified in the simulations for the results in rows 1 and 2, therefore no variances of present value of total cost resulted.<sup>28/</sup> However, row 3 shows the result of including a standard deviation about the linear productivity-density regression for  $S_A$ . In this case the mean present value of total cost is the same as in row 1, but the variance of present value of total cost is nonzero.

---

<sup>28/</sup> The slope for harvesting system  $S_L$  was taken from a linear regression fitted to data adapted from the Bérilage (1966) study.

Row 4 of Table 3 shows the effects of assuming a quadratic productivity-density function for harvesting system  $S_L$ . The shape of this relationship is shown in Figure 7. The standard deviation about the regression was assumed to be zero, yet a variance of the present value of total cost resulted. This is in contrast to row 2 where the variance of present value of total cost was zero for the linear function. This occurs for the linear function because the total time required to pick a given number of trees is the same regardless of the order in which the trees are picked; whereas, for the curvilinear function the total time to pick the same trees is different depending upon the order in which the trees are encountered. It is thus hypothesized that linearity of the productivity-density relationship does not account for the effect of yield variability between trees on the productivity of a harvesting system. It is an empirical question whether a productivity-density relationship is linear or non-linear. The sensitivity tests as conducted with the simulation program served to make this question explicit, and to emphasize the need for more extensive empirical investigation of the nature of the relationship. The sampling formula and procedures were developed using the assumption of a linear relationship. This was done because of lack of evidence of any other, and because linearity was the simpler assumption upon which to base the theoretical derivation of the sample size formula. However, these considerations emphasize the importance of making observations on a system's productivity over a range of fruit density.

In Table 3, rows 5 and 6 show the results for hypothetical quadratic productivity-density functions for an alternative harvesting system  $S_A$  and a conventional ladder system  $S_L$ . These hypothetical curvilinear productivity-density relationships are shown in Figure 7. The slope of  $S_A$  implies that this system becomes technically more efficient at lower fruit density, whereas, the ladder system  $S_L$  becomes technically less efficient at lower fruit density. Using these quadratic functional relationships as input to the computer program, the mean and variance of present value of total cost for each system were calculated on the basis of 50 observations of present value of total cost. Given the means and variances of present value of total cost for actions  $S_A$  and  $S_L$ , a harvest system could be selected upon the decision criterion of either minimization of present value of total cost or minimization of expected disutility for present value of total cost.

Table 8. Summary of results from the sensitivity tests for hypothetical situations.

Har-vesting system	Row number	Productivity-density function coefficients		Standard deviation about regression	Number of bins picked $\bar{c}$	Mean present value of total cost (n=50)	Variance of present value of total cost	Standard deviation of present value of total cost
		$\bar{a}$	$\bar{b}$					
S <sub>A</sub>	1	0.0	58.81	0.0	100.88	\$2,336.38	0.0	0.0
S <sub>L</sub>	2	0.0	67.29	0.0	100.88	1,324.23	0.0	0.0
S <sub>A</sub>	3	0.0	58.81	10.16	100.88	2,335.81	12.79	3.58
S <sub>L</sub>	4	59.167	0.833	0.0	100.88	1,306.96	43.20	6.57
S <sub>A</sub>	5	55.625	-0.625	10.16	100.88	2,082.25	46.21	6.80
S <sub>L</sub>	6	59.167	0.833	10.16	100.88	1,306.46	61.81	7.86

a/ Harvesting systems S<sub>A</sub> and S<sub>L</sub> correspond respectively to a hypothetical mechanical harvesting system and a conventional ladder system.

b/ The productivity-density relationships for the harvesting systems in rows 1, 2, and 3 are described by the general linear regression equation, times per bin = a+b (trees per bin) + error. The productivity-density relationships for the harvesting systems in rows 4, 5, and 6 are described by the general nonlinear regression equation, time per bin = a (trees per bin) + b (trees per bin)<sup>2</sup> + error.

c/ Use of the same initial random number for selection of trees from the specified orchard allows the trees to be picked in the same order for each simulation. This permits a standardized comparison between systems, and hence each system in effect picks not only the same number of trees, but the same total yield.

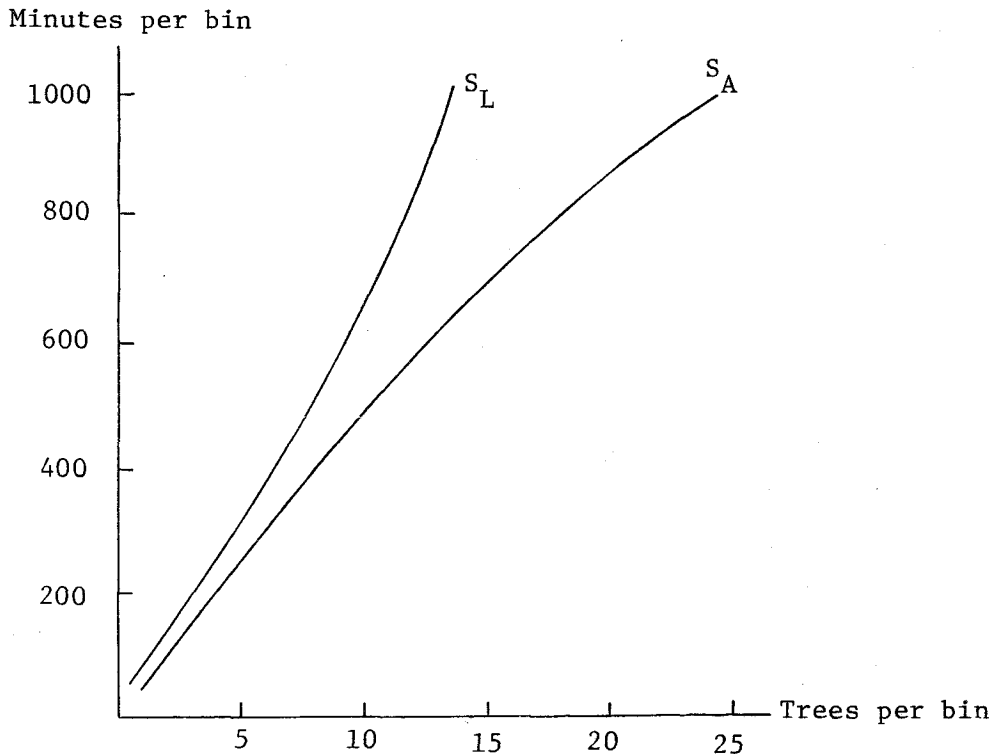


Figure 7. Hypothetical productivity-density relationships for harvesting systems  $S_A$  and  $S_L$ .<sup>a/</sup>

<sup>a/</sup> The curve representing the productivity-density relationship for system  $S_A$  can be expressed as time per bin =  $55.625$  (trees per bin) -  $0.625$  (trees per bin)<sup>2</sup>. The curve representing the productivity-density relationship for system  $S_L$  can be expressed as time per bin =  $59.167$  (trees per bin) +  $0.833$  (trees per bin)<sup>2</sup>.

A decision maker would choose that action which has the smallest mean present value of total cost if his decision criterion is minimization of cost, i.e., he would choose harvesting system  $S_L$  with its mean of \$1,306.46. However, if his decision criterion is minimization of expected disutility, the E-V analysis is applied. In Figure 8 mean present value of total cost is measured along the horizontal axis and variance of present value of total cost along the vertical axis. The specific mean and variance values for  $S_A$  and  $S_L$  are plotted in Figure 8 from rows 5 and 6 of Table 8.



Variance of present  
Value of total cost

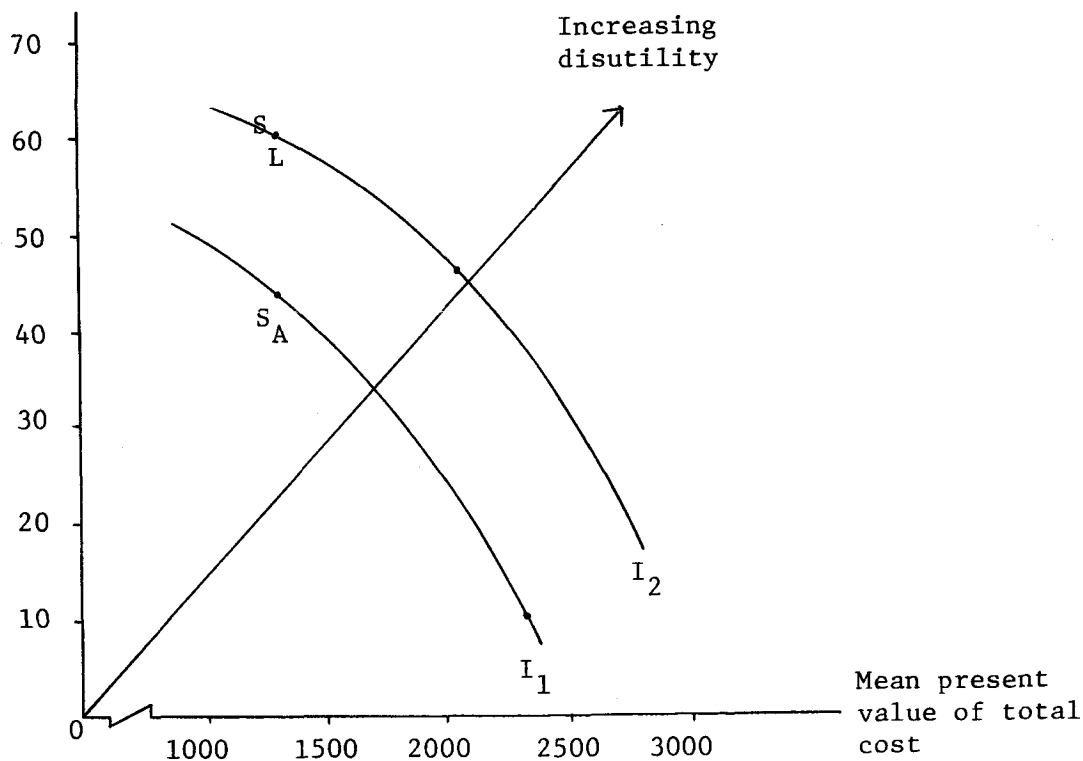


Figure 8. Hypothetical indifference curves and actions  $S_A$  and  $S_L$ .

Levels of utility are shown in indifference curves joining combinations of mean and variance of present value of total cost to which a hypothetical decision maker is indifferent. Because Figure 8 illustrates levels of disutility for cost, the indifference curves are labeled from left to right in order of increasing disutility ( $I_1 > I_2$ ). With this set of indifference curves, the criterion of minimization of disutility results in the choice of the alternative harvesting system  $S_A$  over that chosen by the criterion of minimization of present value of total cost,  $S_L$ . Thus the choice of an apple harvesting system depends ultimately on the individual producer's attitude toward risk.

## BIBLIOGRAPHY

- Barker, M. G. 1960. Mechanization and farm management. In: Mechanization in agriculture, ed. by J. L. Meij. Chicago, Quadrangle Books. p. 98-177.
- Barnes, Ralph M. 1968. Motion and time study. 6th ed. New York, John Wiley and Sons. 799 p.
- Berlage, Arnold G., R. D. Langmo, and G. E. Yost. 1966. Man-positioning principles and their effect on apple harvesting efficiency. Study conducted at Tree Fruit Research Center, Wenatchee, Washington.
- Berlage, Arnold G., Ronald B. Tukey and Gilbert E. Yost. 1967. Modifying tree fruit culture to engineering needs. Paper presented at 1967 annual meeting of the Pacific Northwest Region American Society of Agricultural Engineers, Spokane, Washington, October 18-20, 1967.
- Billard, Jules B. 1970. The revolution in American agriculture. National Geographic 137(2):147-185.
- Burkner, P. F., G. K. Brown, E. G. Vis, and R. C. Rock. 1968-69. Economics and labor advantage of picker-positioning equipment for citrus harvesting. The California Citrograph 54 (1, 2, 3): 25-28, 78-80, 103-110.
- Doud, Lorne J. 1962. Two decades with dwarfs. American Fruit Grower, January 1962, p. 22-24.
- Draper, N. R. and H. Smith, 1966. Applied regression analysis. New York, John Wiley and Sons. 407 p.
- Edwards, Mike W. 1970. Shenandoah, I long to hear you. National Geographic 137(4):554-588.
- Fridley, R. B. and P. A. Adrian. 1966. Mechanical harvesting equipment for deciduous tree fruits. Davis, Calif. 56 p. (University of California. California Agricultural Experiment Station. Bulletin 825).
- Gaston, H. P. and J. H. Levin. 1953. Time and motion studies of apple picking made to determine the possibilities of mechanizing harvest operations. East Lansing, Michigan. 6 p. (Michigan State University. Michigan Agricultural Experiment Station. Quarterly Bulletin, Vol. 36, no. 1, p. 18-23).

- Halter, A. N. and G. W. Dean. 1971. Decisions under uncertainty with research applications. Cincinnati, Smith-Western Publishing Company.
- Heady, Earl O. 1960, Extent and conditions of agricultural mechanization in the United States. In: Mechanization in agriculture, ed. by J. L. Meij. Chicago, Quadrangle Books. p. 63-97.
- Leftwich, Richard H. 1961. The price system and resource allocation. Rev. ed. New York, Holt, Rinehart and Winston. 381 p.
- Maksoudian, Y. Leon. 1969. Probability and statistics with applications. Scranton, International Textbook Company. 416 p.
- McKenzie, D. W. 1969. Cost of harvesting apples: a survey of present methods and possible future developments. Fruit Research Division. Havelock North, New Zealand. 23 p.
- Miller, Davis W. and Martin K. Starr. 1960. Executive decisions and operations research. Englewood Cliffs, New Jersey, Prentice-Hall. 446 p.
- Ricks, Donald J. 1966. Worker productivity in apple picking. East Lansing, Michigan. 24 p. (Michigan State University. Rural Manpower Center, Report No. 1).
- Rudkin, T. H. A sampling procedure to detect alternative actions in a decision model for choosing among apple harvesting systems, M.S. Thesis, Oregon State University, Corvallis, 1971.
- Seamount, Daniel T. 1969. Consistency of picker rates in oranges as a factor in influencing performance on man positioning devices. Paper presented at the 1969 Winter Meeting American Society of Agricultural Engineers, Chicago, Illinois, December 9-12, 1969.
- Simpson, Robert. 1966. Strippers and shakers. Steelways 22(4):6-10.
- Snedecor, George W. and William G. Cochran. 1967. Statistical methods. 6th ed. Ames, Iowa, The Iowa State University Press. 593 p.
- U.S. Dept. of Agriculture. 1969. Agricultural statistics. Washington, D.C., U.S. Government Printing Office. 631 p.

APPENDIX

## APPENDIX

### The Sample Size Formula for Estimating Productivity-Density Relationships and for Detecting a Significant Difference

Developed by Kenneth Burnham

The sample size formula on page 34 is based on a number of assumptions, the main being that the relationship between time per bin and trees per bin is linear and goes through the origin for both the ladder and the machine. Letting  $y$  = time per bin, and  $x$  = trees per bin, we are assuming  $E(y) = xB$ . To distinguish observations made with the ladder from those of the machine, a subscript will be used; e.g.,  $y_L$  = time per bin when the ladder is used. Finally, let  $B_L$  = the slope of the regression line when the ladder is used, and  $B_M$  = the slope of the regression line when the machine is used.

The statistical problem is to test whether  $B_L - B_M \leq d_s$  or  $B_L - B_M > d_s$ , where  $d_s$  is some specified difference between these slopes. To carry out such a test, one must take observations on both methods of picking apples. A natural question to ask is how many observations should be taken to detect the case  $B_L - B_M > d_s$ . The statistician can shed some light on this question but he cannot give a definitive answer because he does not know the true value of  $B_L - B_M$ . The approach to this problem is to first determine a test statistic which would be used if the data were available.

It is best if the same number of observations are taken with both ladder and machine. The full model for the problem is then:

$$y_{Li} = x_{Li} B_L + e_{Li} \quad i = 1, \dots, N$$

$$y_{Mi} = x_{Mi} B_M + e_{Mi} \quad i = 1, \dots, N$$

where  $e_{L1}, \dots, e_{LN}, e_{M1}, \dots, e_{MN}$  are independent normal random variables, all with mean zero and unknown variance  $\sigma^2$ . Symbolically we write  $E \sim N(0, \sigma^2)$ . Note that a total of  $2N$  observations are taken.

The least squares estimates of the parameters  $B_L$  and  $B_M$  are:

$$\hat{B}_L = \frac{\sum_{i=1}^N (x_{Li} y_{Li})}{\sum_{i=1}^N (x_{Li})^2}$$

$$\hat{B}_M = \frac{\sum_{i=1}^N (x_{Mi} y_{Mi})}{\sum_{i=1}^N (x_{Mi})^2}$$

These estimators are independent random variables.

$$\hat{B}_L \sim N \left( B_L, \frac{\sigma^2}{\sum_{i=1}^N (x_{Li})^2} \right)$$

$$\hat{B}_M \sim N \left( B_M, \frac{\sigma^2}{\sum_{i=1}^N (x_{Mi})^2} \right)$$

Consequently  $\hat{B}_L - \hat{B}_M$  is also normally distributed:

$$\hat{B}_L - \hat{B}_M \sim N \left( B_L - B_M, \sigma^2 \left[ \frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2} \right] \right)$$

Let  $B_L - B_M = d_o$ , the true difference between the slopes of the line, then an estimator of  $d_o$  is  $\hat{d}_o = \hat{B}_L - \hat{B}_M$ . Remembering that  $d_s$  is some specified difference between the slopes, we have

$$\hat{d}_o - d_s \sim N \left( d_o - d_s, \sigma^2 \left[ \frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2} \right] \right)$$

Now define the test statistic  $W$ :

$$W = \frac{\hat{d}_o - d_s}{\sigma \sqrt{\frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2}}}$$

we have

$$E(W) = \frac{d_o - d_s}{\sigma \sqrt{\frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2}}}$$

and  $W \sim N(E(W), 1)$ .

The hypothesis to be tested can be stated as

$$H_0 : d_o - d_s = 0 \quad \text{versus} \quad H_1 : d_o - d_s = \delta > 0$$

The test procedure is to reject  $H_0$  if  $W \geq Z_\alpha$ , where  $Z_\alpha$  is the upper  $1 - \alpha$  point of the  $N(0, 1)$  distribution. When  $H_0$  is true  $P[W \geq Z_\alpha] = \alpha$ .

If  $H_1$  is true then  $W \sim N(E(W), 1)$  where  $E(W) \neq 0$ . In this case we want to reject  $H_0$ . It is reasonable, therefore, to ask what is the probability that  $W \geq Z_\alpha$ . Because

$$W \sim N(E(W), 1) \quad ,$$

we have, for all values of  $d_0 - d_s$ ,

$$W - E(W) \sim N(0, 1) \quad .$$

Also,  $W \geq Z_\alpha$  occurs if and only if  $W - E(W) \geq Z_\alpha - E(W)$ . Let  $W^* = W - E(W)$  and we have

$$\Pr [ W \geq Z_\alpha ] = \Pr [ W^* \geq Z_\alpha - E(W) ] ,$$

where  $W^* \sim N(0, 1)$ , for all values of  $d_0 - d_s$ .

This probability of rejecting  $H_0$  when it is false is called the power of the test. For any fixed value of  $d_0 - d_s = \delta > 0$  the larger  $N$  is, the more powerful the test. The statistician can now state the problem of sample size in the following terms. If the true value of  $d_0 - d_s = \delta > 0$ , then how large a sample must we have to achieve probability  $\beta$  (e.g.,  $\beta = .90$ ) of rejecting  $H_0$  when we use an  $\alpha$ -level test?

The answer is, we want to choose  $N$  such that

$$\Pr [ W \geq Z_\alpha ] = \Pr [ W^* \geq Z_\alpha - E(W) ] = \beta .$$

Tables of the  $N(0, 1)$  distribution are readily available, so it is easy to find the number  $Z_\beta$  such that  $\Pr [ W^* \geq Z_\beta ] = \beta$ . To find the desired sample size, we equate

$$Z_\beta = Z_\alpha - E(W) \quad ,$$

and solve for  $N$  after appropriate substitutions are made as follows:

$$E(W) = Z_\alpha - Z_\beta$$



$$\frac{d_o - d_s}{\sigma \sqrt{\frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2}}} = Z_\alpha - Z_\beta$$

$$\frac{1}{\frac{1}{\sum_{i=1}^N (x_{Li})^2} + \frac{1}{\sum_{i=1}^N (x_{Mi})^2}} = \frac{\sigma^2 (Z_\alpha - Z_\beta)^2}{(d_o - d_s)^2}$$

In this equation,  $Z_\alpha$  and  $Z_\beta$  are constants, easily determined once  $\alpha$  and  $\beta$  are given.

The value of  $\delta = d_o - d_s$  must be specified by the investigator; but  $\sigma^2$  and the  $x$ 's are unknowns which must be estimated. Both machine and ladder will be tested on the same population of trees, therefore we may assume:

$$E((x_M)^2) = E((x_L)^2) = E(x^2),$$

and

$$\frac{1}{N} \sum_{i=1}^N (x_{Li})^2 = E(x^2) = \frac{1}{N} \sum_{i=1}^N (x_{Mi})^2.$$

Substituting in the main equation above, we get

$$\frac{N E(x^2)}{2} = \frac{\sigma^2 (Z_\alpha - Z_\beta)^2}{(d_o - d_s)^2}, \text{ and hence}$$

$$N = \frac{2\sigma^2 (Z_\alpha - Z_\beta)^2}{E(x^2) (d_o - d_s)^2}$$

This formula gives the desired sample size to have probability  $\beta$  of rejecting  $H_0$  with an  $\alpha$ -level test, when the true difference in slopes is  $d_o - d_s > 0$ . The actual use of the sample size formula requires estimates of  $\sigma^2$  and  $E(x^2)$ , hence the resultant value of  $N$  is only an approximation of the desired sample size.

Note  $Z_{.05} = 1.645$   
 $Z_{.90} = -1.28$