AN APPLICATION OF LINEAR REGRESSION AND CORRELATION ANALYSIS TO THE POULTRY PROCESSING INDUSTRY

By

ROY C. FERGUSON III

Bachelor of Science

Oklahoma State University

Stillwater, Oklahoma

1968

Submitted to the Faculty of the College
of Business Administration
of Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF BUSINESS ADMINISTRATION
May, 1970

PREFACE

The poultry processing industry has changed markedly in the past decade. The large volume processing plant has emerged as the dominant force in the industry replacing the small, family owned plants which were prevalent for many years.

Because processing plants are becoming larger, investment in such plants is also increasing. For example, a plant capable of turning out 15 million pounds of product annually requires an investment of two million dollars. As a result, detailed knowledge of the cost function should be present.

However, from a survey of six modern plants, it is evident that only limited knowledge exists with respect to the cost function. Plant managers are unable to discuss in detail what variables most influence the cost function.

Additional study of the cost function is thus warranted. Any study conducted should be analytical in nature rather than intuitive as has been the case too often in the past. One analytical technique that may possess potential with respect to the poultry processing industry is linear regression and correlation analysis. The determination of whether a potential does in fact exist is the purpose of this study.

The data for this study was taken from one processing plant. Because of the wishes expressed by the manager of the plant under consideration, no direct mention is made either of the particular plant or the state within which it is located.

Gratitude is expressed to the plant manager who furnished the data for this study. In addition, special indebtedness is acknowledged to Dr. Kent Mingo without whose constant help and encouragement this study could not have been completed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background of the Study

Throughout the poultry industry in the United States, the need for large, mechanized processing plants is recognized because of the economies of scale which can be realized from such plants. Quickly vanishing is the breed of small plants which up to even ten years ago was a familiar item. Taking a southeastern state as an example, one may find today only three commercial processing plants in operation. One plant is housed in what originally was a peanut mill; thus, it does not resemble a modern plant in its operational aspects. A second plant is functionally obsolete. Built in 1939, its equipment is aged and its methods impractical. In comparison, the third plant is relatively new. It was constructed in 1967 specifically for turkey processing and employs 165 people. Although this plant is certainly not the largest modern poultry processing plant in the United States, it is far from the smallest. Assuming a weekly six-day schedule with one eight-hour shift daily, the plant is capable of turning out forty million pounds of finished product annually.

However, because of the general shortage of turkeys in the state, the plant has been forced this year, as in past years, to haul in turkeys from as distant a radius as 800 miles to increment the state's production. Even so, the plant has had volume to warrant only six months of operation this year. The other six months the plant has remained closed. As a result, the full potential of the operation has not been realized, and profit margins are very small. It, therefore, becomes very important that the plant management pay extremely close attention to all possible factors which may influence costs.

Purpose of the Study

One method which is used to analyze the cost function in other industries is that of linear regression and correlation analysis. The purpose of this study is to determine whether or not a potential exists for the application of such analysis to the poultry processing industry. Regression and correlation analysis, if applicable to processing costs, would be very useful to management because statistical methods would be implemented as part of regression and correlation techniques. The whole cost study would then become more analytical and detailed than is currently the practice. Better cost information in turn could only lead to more intelligent decision making by management.

Methodology

Linear regression and correlation analysis allows one to isolate variables and determine their individual importance with respect to cost. In addition, combinations of variables may be taken to determine the joint influence of several variables on cost. With the aid of a multiple linear regression program furnished by the Computer Science Department of Oklahoma State University, a sample of cost data has been extracted from the newest processing plant in which twenty variables have been utilized as depicted in Table I.

One variable is chosen to be the dependent variable. The other nineteen variables are designated as independent variables. The dependent variable, cost, is determined by the relationships existing among the nineteen independent variables.

In this study, several combinations of the independent variables were taken to determine the relationships between them and the dependent variable. The discussion of the results will follow later in the data analysis chapter of this paper. It suffices to say that major differences in importance among the independent variables were found in the computer program analysis.

Hypothesis of the Study

The major hypothesis of this study is that the cost

TABLE I

LIST OF VARIABLES USED IN STUDY

Dependent Variable

(1) Average Cost per Pound

Independent Variables*

(2)	Volume of Hens-lbs.	(12)	Sanitation
(3)	Volume of Toms-lbs.	(13)	Supervision
(4)	Number of Hens	(14)	Plant Indirect Labor
(5)	Number of Toms	(15)	Maintenance
(6)	Kill and Pick-lbs. per man hour	(16)	Refrigeration
(7)	Eviscerate-lbs. per man hour	(17)	Turnover
(8)	Pack-lbs. per man hour	(18)	Average Days Run per Period
(9)	Box-lbs. per man hour	(19)	Breed of Turkey(s) Processed
(10)	Average Eviscerated Weight	(20)	Distance Hauled
(11)	Warehouse and Shipping		

*In order to avoid any confusion which might arise as a result of a lack of understanding with respect to the meaning of certain independent variables, seven definitions are in order:

(1) Hens - A female turkey.
(2) Tom - A male turkey.
(3) Kill and Pick - The process in which the turkey is killed and the feathers removed.
(4) Eviscerate - The process in which the body cavity of the turkey is cut open and the internal organs removed.
(5) Pack - The process of placing the processed turkey in a vacuum sealed bag.
(6) Box - The process of placing the vacuum sealed bag containing the processed turkey into a container suitable for shipping.
(7) Sanitation - The process of keeping the plant hygienically clean.

function is such that one may ascertain the importance of some of the listed independent variables in determining the dependent variable with a stated statistical degree of confidence.

It is recognized that the scope of this hypothesis may seem to be very broad. However, after a search through past applied uses of regression and correlation analysis, no results were found of any previous direct applications to the poultry processing industry. One should not be misled to accept the assumption that regression and correlation methods are unapplicable to the poultry processing industry. As will be presented in the literature review, there are cases where regression and correlation analysis has been applied to industries that are very similar to the poultry processing area.

Limitations

Because of both time and monetary constraints involved, certain limitations are present in this study. For instance, only a limited sample of data from one plant was obtained. Ideally, one would have preferred to collect data from several plants over a longer time period than was possible. Such a process would have permitted the same sample to be increased with respect to both firms and cost periods involved and, thus, would have more accurately represented a normal distribution. Additionally, it would have been preferred that one be able to draw from previous studies directly applied to the poultry processing industry.

Overview

Although this study does have certain described limitations, it still merits consideration. Again, this paper should not be considered to be an all-encompassing empirical study, for as has been mentioned, it is not. But rather, it should be thought of as an exploratory one, seeking to uncover relationships that are operational in the production function of the industry. It is hoped that as a result of this paper, individuals, particularly in the processing industry, will make refinements to further apply regression and correlation analysis to improve the efficiency of the industry.

Brief discussion has been presented in this chapter with respect to the background, purpose, methodology, hypothesis, and limitations involved in this study. Chapter II contains a survey of the literature pertinent to regression and correlation analysis. Chapter III holds the detailed methodology which is implemented in this paper as well as a thorough description of the limitations. Chapter IV presents the data analysis which is the combined collection of important results found. Chapter V contains a brief summary of the paper as well as implications and future regions of possibility for regression and correlation analysis.

CHAPTER II

LITERATURE REVIEW

This chapter will present the relevant literature which justifies a statistical cost analysis of the poultry processing industry. Primarily, the discussion will revolve around past applications of regression and correlation analysis to the electrical and sugar refining industries.

Chapter I has suggested that a study of poultry processing cost functions is needed. Furthermore, it has stated that no direct applications of regression and correlation analysis have been made within the poultry industry. However, there have been previous studies which, while not directly connected to the poultry processing industry, may still be discussed because of the same applications involved.

A Study From the Electrical Industry

Specifically, one may look to the writings of Dr. John Johnston, professor of econometrics at the University of Manchester. Johnston divides his studies of cost functions into two categories:

(1) short period in which the firm's actions are subject to the constraint that plant,

equipment, and buildings cannot be changed quickly in amount, and

(2) long period in which no such constraint is present.[1]

It is from the short period information that this study will draw since Johnston's hypotheses on short run costs involve the assumption that the firm's activities are constrained by some fixed limit.

According to Johnston, one should ideally use a series of paired observations on costs and output which satisfy the following conditions:

(1) The basic time period for each pair of observations should be one in which the observed output was achieved by a uniform rate of production within the period. As an example, one would not wish to choose as his period of basic production a week if there were substantial variations in daily production within each week.[2]

This study satisfies condition (1) in that daily production did not fluctuate greatly within each weekly period.

(2) The observations on cost and output should be properly paired in the sense that the cost figure is directly associated with the output figure.

This study satisfies condition (2) in that all accounting data for weekly periods utilized the same time period. As an example, sanitation costs for week two were based on hours worked in week two rather than week one.

(3) Ideally, a wide spread of output observations

---

[1]John Johnston, Statistical Cost Analysis (New York, 1960), p. 46.

[2]Ibid., pp. 26-27.

> is desired so that cost behavior can be observed at widely differing rates of output.

This study satisfies condition (3) in that greatly differing rates of output were observed. Volume in pounds ranged from 162,550 pounds to 677,000 pounds per week. In addition, the plant varied in weekly operations from two days to six days of production.

> (4) It is necessary to keep the experimental data uncontaminated by the influence of factors extraneous to the cost-output relationship itself.

Condition (4) is also satisfied in that no price changes occurred which the plant manger felt were significant enough to alter the cost function. In addition, no new technical knowledge became available to the manager during the period of observation which would have allowed him to reduce costs.

Johnston, in his short run example, illustrates an analysis derived from electricity generation. He chooses to draw his sample data from seventeen firms. The current study is different from that of Johnston's. Because of time and monetary considerations as well as a lack of cooperation from processing managers who might have supplied information, only one plant is considered in this paper. However, the effectiveness of this study with respect to the one plant is not impaired; rather, one must realize that it is the scope with respect to all plants which is limited. If one is willing to assume that the plant under consideration is indeed representative of modern poultry processing facilities, then the entire scope is not even seriously affected.

A Study From the Sugar Refining Industry

Philip Lyle contends that the methods of regression and correlation analysis are being applied to all types of industrial projects with increasing frequency. In particular, he points to the sugar industry where a large tonnage of relatively uniform bulk material is turned out with comparatively few specialties involved.[3] Lyle argues that in the practical applications of regression and correlation analysis to sugar refineries, he has found managers who are now making regular use of these methods.

While a poultry processing plant certainly is different from a sugar refinery in many areas, it nevertheless has certain common aspects. For instance, the poultry product is a rather uniform entity, and as such, one may wish to define it in tonnage as is the common practice. The sugar industry also produces a uniform product and defines its finished product, refined sugar, in tonnage refined.

In addition, few specialties are involved in the poultry processing industry just as is the case in the sugar refining process. It is true that individuals on the

---

[3]Philip Lyle, Regression Analysis of Production Costs and Factory Operations (New York, 1957), pp. 107-111.

poultry processing line perform particular tasks. However, employees may interchange with little or no loss of efficiency involved given a reasonable time period to adjust to the new task. The important point is that the whole process is fairly routine just as is the case in a sugar refinery.

As this study is designed to be exploratory in nature, it is suggested that one purpose will be accomplished if individual processing managers recognize the potential applications that presently lie unused and seek professional advised as to how these applications may be implemented.

It is clear that there are facets of the electricity and sugar refining cases which are similar in nature to the poultry processing situation discussed in this paper. The short run example as illustrated by Johnston and the uniform product as described by Lyle mesh strongly into the framework of this paper. The discussion will now turn to the methodology implemented in this paper as well as the limitations which are present in the study.

CHAPTER III

METHODOLOGY

This chapter contains reasons for the methods used in:

(1) collection of information

(2) selection of linear regression program

(3) definition of terms.

In addition, each limitation which was encountered in this study is listed, and explanations are given as to why the limitations did place constraints on the analysis.

Collection of Information

In the initial stages of this study, a selection of more than one poultry processing plant was planned in order to observe the difference in efficiencies existing among processing entities. However, after discussions with six plant managers, it became evident that little cooperation could be obtained in the collection of cost data. Only one manager agreed to furnish the necessary information, and for that reason, one plant is the focal point of this study.

Objective accounting records from the selected plant supplied data which allowed the selection of the twenty variables listed in Table I. All twenty variables are backed by explicit records for each weekly period involved.

Selection of the Linear Regression Program

Having determined the availability of objective records, the next step was to determine in exactly what form or forms this data could be used so that the cost function could be analyzed properly. It was found that the data was in such a form that it could be implemented into the library linear regression program supplied by Oklahoma State University.

With the selection of the specific program determined, a series of questions were developed and put into the form of a questionnaire in order to gather the necessary information which could be used in the computer analysis (see Appendix A). Most of these questions were objective in nature. However, some questions were subjective. The purpose of the subjective questions was to probe the intuitive feelings of the plant manager to determine what he believed to be the important behavioral variables relating to cost. For example, the manager was asked if he thought sanitation expense within the plant affected cost significantly. Although he thought sanitation to be rather unimportant, the data analysis proves otherwise. This fact alone demonstrates that the cost function is more complicated than the plant manager had thought.

Definition of Terms

Since the terms regression and correlation have been used and will continue to be used throughout the length of this paper, it is necessary to define them as well as the uses of a few other terms in the context of their meaning here. The five terms immediately following are defined by Philip Lyle:

(1) Regression Equation - an equation by which we estimate the value of a dependent variable for given values of the independent variable.

(2) Multiple Regression Equation - a regression equation with two or more independent variables.

(3) Regression Line - an equation by which we estimate the value of a dependent variable for given values of the independent variables.

(4) Correlation - the association between two variables, one of which may be a linear function of several different variables.

(5) Correlation Coefficient - A convenient measure of correlation lying in the ranges 0, for no association whatever, to ± 1 for perfect linear correlation. The positive range covers cases where the variables tend to increase or decrease together, the negative where one tends to increase or the other decrease in value. The term "Multiple Correlation Coefficient" is used when the association being measured is between a dependent variable and two or more independent variables acting together, but is identical with the simple correlation coefficient between the dependent variable and its estimate by a regression equation.[1]

---

[1]Philip Lyle, <u>Regression Analysis of Production Costs and Factory Operations</u> (New York, 1957), pp. 198-202.

Although the five preceding terms are defined in the reference of the given study, Mordecai Ezekiel contends that there are three additional constants which summarize nearly all that a correlation analysis reveals:

> (1) The standard error of estimate shows how nearly the estimated values agree with the values actually observed for the variable being estimated. This coefficient is stated in the same units as the dependent variable, and its size can be compared directly with those values.
>
> (2) The coefficient of determination shows what proportion of the variance in the values of the dependent variable can be explained by or estimated from the concomitant variation in the values of the independent variable. The coefficient of determination is the square of the correlation coefficient.
>
> (3) The coefficient of simple linear regression measures the slope of the regression line. That is, it shows the average number of units increase or decrease in the dependent variable which occur with each increase of a specified unit in the independent variable.[2]

It is important to note that although the three terms defined by Ezekiel measure certain aspects of the relationship among variables, it does not follow that all three terms will vary together. Similarly, a problem which shows a high coefficient of determination will not necessarily show a high regression coefficient or a low standard error of estimate. That is because they measure different aspects of the relation.[3]

---

[2]Mordecai Ezekiel, Methods of Correlation Analysis (London, 1941), p. 159.

[3]Ibid., p. 160.

Limitations of the Study

It is important to remember with regard to this study that the correlation coefficient, regression coefficient, and standard error of estimate will often not vary in proportion. In fact, most of the regression coefficients are very small, and some are even zero. The smallness is due to the fact that the dependent variable, cost per pound, is expressed in low figures - $.015 to $.030 per pound. Therefore, the slope of the regression line will be accordingly small. Due to rounding procedures in the computer program, values of zero for the regression coefficient were sometimes obtained. This was a recognized limitation, but it could have been eliminated had the computer program been designed to carry out cost figures to more than five decimal places.

One particular limitation that did occur as a result of the small sample concerned the numbers of independent variables which could be combined at any one time. It was hoped originally that one would be able to make combinations of the independent variables in any fashion desired. However, it was discovered that the library program requires that the number of independent variables combined in any one sequence be less than the number of observations. Since the number of observations in this study was equal to eight, a theoretical maximum of seven was then imposed as a constraint within which it was necessary to work.

Furthermore, it was found that because of the F tests involved (statistical tests of significance), it was infeasible to use a combination of more than three independent variables at a time. The exact reason for this limitation was traced to the small sample size. As a sample decreases in size, it becomes necessary to decrease the number of independent variables used in combination. Otherwise, any results obtained are likely to be statistically insignificant.

With the forementioned constraints in mind, different sets of combinations of independent variables were designed. All of the independent variables were tested singlely against the dependent variable to determine the individual correlation coefficients and the levels of significance involved. After these results were evaluated, combinations of two and three independent variables were taken to determine what total effects they had on the dependent cost variable as opposed to just single independent versus dependent variable trials. In total, more than 65 different trials were conducted. As one might expect, the results varied from highly significant to relatively insignificant, depending on the exact combination of variables.

Such a range of significance had been desired in the initial stages of the study. One purpose, after all, had been to locate and isolate variables which were of importance relating to cost. After location of the important variables was made, determination of their exact association with the cost function followed. As the results indicate in the data analysis chapter, answers were found which were almost totally unexpected relating to the supposedly insignificant variables.

Summary

In this chapter, the methodology involved has been discussed. The methods of linear regression and correlation analysis have been defined as they are implemented in this study. Important statistical terms, such as the F test, regression coefficient, and coefficient of determination, which often carry unclear meanings have also been defined. Chapter IV, data analysis, will concentrate on the actual results and the implications arising from the results.

CHAPTER IV

DATA ANALYSIS

This chapter will discuss the relationships which were found to exist among the dependent variable, cost, and several of the independent variables. The independent variables. The independent variables which warrant the most discussion, based on their correlation coefficients and coefficients of determination as well as F tests and implications which may be drawn, are volume of hens and tons, numbers of hens and toms, kill and pick, eviscerate, sanitation, supervision, indirect labor, turnover, average days run per period, breed of turkey processed, and distance hauled. As one can readily see, several variables merit consideration. The exclusion in the discussion of some variables does not necessarily mean that they are totally unimportant. Rather, they are excluded only because they are not so important as the forementioned variables.

Existence of an Inverse Relationship Among the Dependent and Independent Variables

Inasmuch as the manager of the plant studied had felt that volume was perhaps the most important factor influencing cost, one might expect the matrix of correlation coefficients of the independent variables to the dependent variable to be inversely related. This expectation stems from the general reason that as volume decreases, plant efficiency also decreases because employees lose familiarity with their tasks. In addition, the influence of fixed costs to total costs will be greater when volume is low than when it is high.

Table II gives the single correlation matrix along with the associated F tests for the variables used in this study. In Table II, an F value of 3.77 is necessary for statistical significance at the .10 level. Most variables then do not carry such significance. However, this could have been remedied had the sample size been larger. In fact, a sample of only twenty observations is normally needed to insure approximate conformation to the normal distribution.[1]

To better understand what the F test means, one should look at variable 12, sanitation, when correlated with variable one, cost per pound. Using the F table, one may determine that the given correlation coefficient of -0.77 is significant at the .02 level. That is, only .02 of the time would one expect to obtain a correlation as large as -0.77 by chance.

One may determine that all of the correlation coefficient signs in Table II are negative. One would expect this

---

[1]Mordecai Ezekiel and Karl A. Fox, Methods of Correlation and Regression Analysis (New York, 1959), p. 10.

TABLE II

CORRELATION MATRIX WITH F TESTS USING THE POULTRY PROCESSING VARIABLES

		(1) Cost per Pound				(1) Cost per Pound	
		Correl. Coeff.	F Test			Correl. Coeff.	F Test
(2)	Vol. of Hens-lbs.	-0.32	0.72	(12)	Sanitation	-0.77	9.03
(3)	Vol. of Tons-lbs.	-0.41	1.22	(13)	Supervision	-0.01	0.001
(4)	No. of Hens	-0.35	0.98	(14)	Plant Indir Labor	-0.81	11.52
(5)	No. of Toms	-0.49	1.91	(15)	Maintenance	-0.12	0.10
(6)	Kill and Pick	-0.61	3.68	(16)	Refrigeration	-0.36	0.89
(7)	Eviscerate	-0.11	0.08	(17)	Turnover	-0.48	1.80
(8)	Pack	-0.50	2.05	(18)	Ave. Days Run per Period	-0.23	0.36
(9)	Box	-0.62	3.94	(19)	Breed of Turkeys Processed	-0.36	0.92
(10)	Ave. Evisc. Wt.	-0.27	0.48	(20)	Distance Turkeys Hauled	-0.55	2.70
(11)	Warehouse and Shipping	-0.51	2.17				

*A value of 3.77 is required for significance at the .10 level.

result for two specific reasons.

First, certain costs are considered fixed. At a low output of tonnage, these fixed costs would be allotted over relatively few pounds of product. Thus, cost per pound would tend to be higher at low tonnage than at high tonnage.

As an example, one might consider depreciation. Poultry processing equipment is such that no or virtually no use does not conserve the equipment. In fact, because of the production line involved (i.e., the stream of equipment where product flows from start to finish as in an automobile assembly plant), the corporate owners of this plant would rather the entity be in uniform operation. Because of this, depreciation is considered to be a fixed cost. If tonnage increases threefold, depreciation costs remain the same. The significant change as far as this study is concerned is that a threefold increase in tonnage causes depreciation costs per pound to decrease to one-third of the previous amount.

Further examples of fixed costs could be given. Some maintenance, refrigeration, and supervision are necessary regardless of the level of operation, unless of course the plant is completely shut down for an indefinite period of time. However, the depreciation example should suffice in this instance as the other three examples of allotted fixed costs which have been mentioned are very similar in nature.

Secondly, although other costs are considered to be variable, at very low levels of output they tend to be higher per pound than at high levels of output. In the strictest accounting sense, one might question whether or not these costs are actually totally variable.

According to the plant manager, the skill levels of employees is not as high at low levels of output as at high levels. Thus, a learning curve appears to be present. It is not the intent of this study to determine the precise shape of the curve, but it is present. As higher outputs of tonnage occur, employees become more familiar with their job tasks. Line speed may be increased and, as a result, more pounds per man-hour may be processed. Again, cost per pound may be expected to decrease although not as sharply as in the case of fixed costs.

Results From Combinations of the Independent Variables

Since some 65 combinations of independent to dependent variable trials were conducted, 65 different results were obtained. However, only the results which were felt to be the most important will be discussed. It should be said that some very unexpected findings occurred, and the discussion will now turn to the analysis of the important findings.

The plant manager felt that volume was the most important factor influencing the cost function. Indeed it was because as volume increased, cost per pound decreased due to, among other things, the affect of fixed cost outlays and the so-called learning curve affect.

The amount of hens processed in the eight weeks of data concerned was far less than the amount of toms. Specifically, an average of 7,080 hens were processed each week while the average figure on toms was 16,345. One might expect then that the output for toms would have a higher correlation and, therefore, coefficient of determination than the output for hens.

The data indicates that the correlation for number of hens processed to the dependent variable was .34. In comparison, the data for toms indicates a correlation of .49. The multiple correlation coefficient for the two variables is .78. This in turn gives a coefficient of determination of approximately .60. In addition, the F value is 3.00 which allows the results to be statistically significant at the .10 level. Because of rounding procedures, the regression coefficients for the regression equation involving hens and toms were computed as values of zero. The rounding was caused simply by the particular computer program that carried out numbers to five decimal places. Had the program been designed to use ten decimal places, positive regression coefficients could have appeared.

Using the number of hens and toms to compute estimated cost per pound, results were found that demonstrated a close relationship between estimated and actual cost per pound. Table III gives this relationship.

TABLE III

DEVIATION OF ESTIMATED FROM ACTUAL COST PER POUND USING THE VARIABLES OF NUMBER OF HENS AND TOMS

Case No.	Actual Costs	Estimated Costs	Residual	% Difference
1	$0.01930	$0.02055	$+0.00125	6.00%
2	0.02430	0.02221	-0.00209	10.00%
3	0.02110	0.02109	-0.00001	0.05%
4	0.02110	0.02202	+0.00092	0.40%
5	0.02140	0.02025	-0.00115	5.00%
6	0.02520	0.02544	+0.00024	0.10%
7	0.02800	0.02758	-0.00042	2.00%
8	0.02210	0.02337	+0.00127	0.50%

As can be seen, the greatest residual involved for any one week is $.002, or expressed in percentage terms 10%. Four residuals are less than 1%. The residuals for the variables of number of hens and toms are among the lowest of any variables used in the study. As a predictor of cost, one would, therefore, want to include the two forementioned variables before other variables if only limited information were available.

One might also observe that the residuals involved are evenly divided among positive and negative values. Such a pattern is desirable in that no continual overage or underage is given. As a result, one may plot the estimated values on a scatter diagram and then draw a line to estimate the least squares of the diagram. It is noted that the use of a scatter diagram gives only a rough approximation of the regression relationship. Figure 1 illustrates the use of a scatter diagram to estimate the regression line.

Some interpretation of the scatter diagram is necessary. One notices that the predicted line slopes in an upward direction. This is due mainly to the fact that volume was at its peak in week one. Similarly, volume was at its lowest near the end of the cycle, specifically in week seven. Therefore, one would expect the predicted cost line to slope upward in this instance.

Analysis of the Variables Contained on the Processing Line

The processing line, as has been shown, may be compared to an automobile assembly line in that both processes are continual ones. In discussions with the plant manager, it was discovered that the initial stage of the process, or what is commonly termed "kill and pick", was thought to be very important.

Combining the independent variables of number of hens and toms with the independent variable of "kill and pick", a multiple correlation coefficient of .90 resulted. Accordingly, the coefficient of determination was .81. The F test gave a value of 6.11 which is significant at the .08 level.

The highest multiple correlation was obtained when a combination of the following three variables was taken:

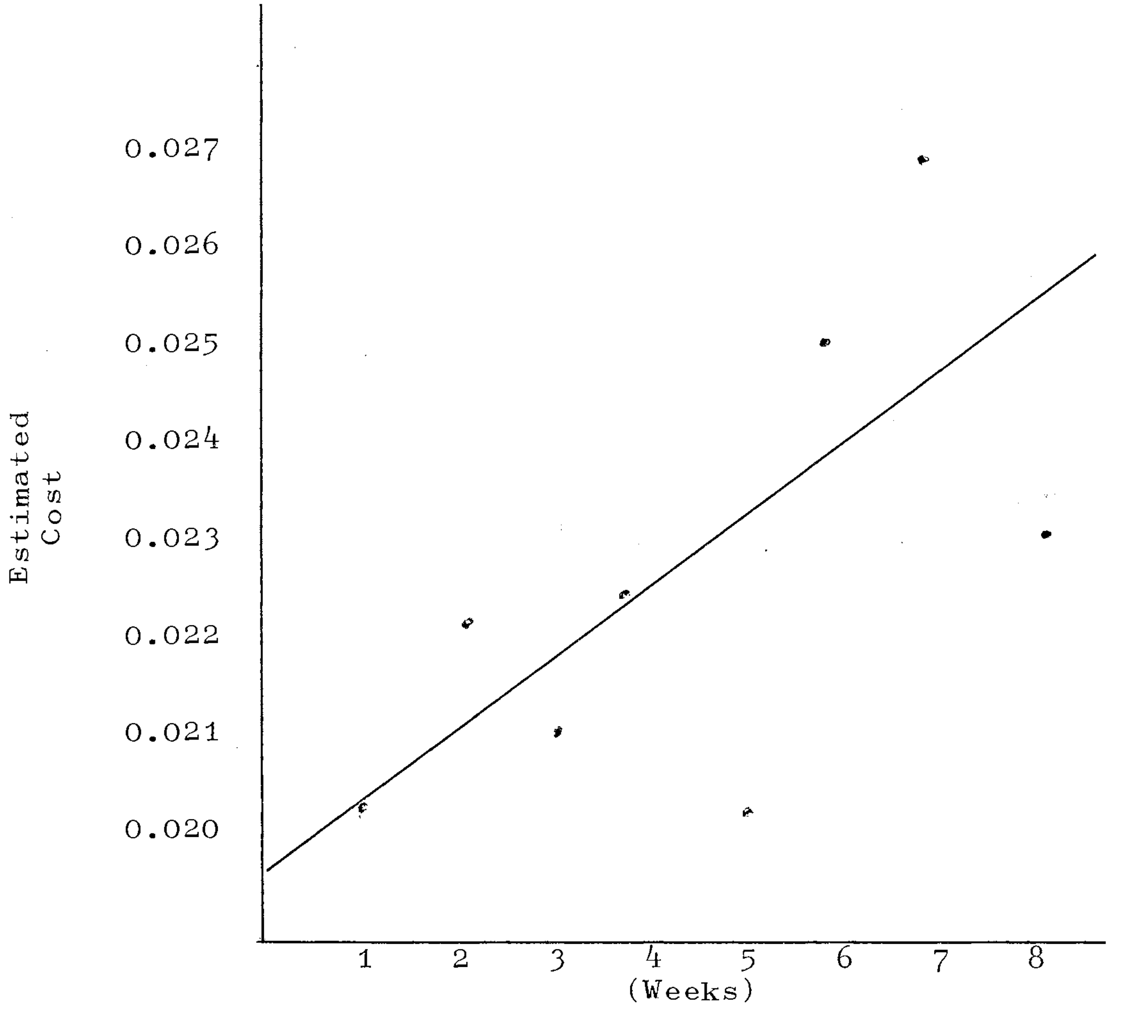

Figure 1. Illustration of the Use of a Scatter Diagram to Estimate the Cost Function

(1) Pounds of toms

(2) Number of toms

(3) Eviscerate - Eviscerate is the stage in the process which immediately follows "kill and pick".

Implementing the forementioned variables, a multiple correlation coefficient of .96 occurred along with a coefficient of determination of .92. The F test gave a significance of .02. The results were especially interesting because the variable eviscerate by itself showed a correlation of only .11 which was the second lowest single correlation coefficient for any one independent variable. However, when the variable was combined with both the number and pounds of toms processed, it became statistically meaningful. One possible explanation for this is that a greater number of pounds of toms may be processed hourly than of hens. One might argue though, and quite correctly so, that a greater number of hens may be processed hourly. One must note, though, that the sum total of hens over toms which may be eviscerated hourly is not enough to offset the greater volume in pounds that may be obtained as a result of processing toms. The reason for this is that a certain amount of labor is required to "cut open" the body cavity of a turkey. The removal of internal organs follows. As such, a hen does not take so long to process as a tom because hens are naturally smaller. However, the time saved by processing a hen is not proportionate to the decrease in body weight, and, therefore, more pounds of toms may be processed hourly than of hens.

The traditional thought within the poultry industry is that supervision is a very important variable as relates to cost. The argument usually put forth is that up to a relatively high point, more supervision will result in lower cost because employees will perform better. In the results of this study, very little support was given to this idea. Supervision when correlated against the dependent variable, cost, gave a correlation coefficient of only .01, the lowest coefficient involved in the study. As a result, the coefficient of determination was .0001, a very low figure. Thus, only .0001 of the variance in cost per pound could be explained by supervision. One might theorize that the manager's job in this case is distinctly repetitive. Once the job is learned, the person involved need only follow established rules. In other words, a high level of skill is not needed.

Presence of a Learning Curve

If a learning curve effect is indeed present, employees should become more efficient up to a certain point as they work more on their assigned tasks. Substantiation of this contention will now be given.

The two variables which would most influence the learning curve effect are turnover and average days run per period. If the learning curve exists, one would expect that as turnover decreased and the average days run per period increased, cost per pound would decrease.

In this study, turnover and average days run per period were combined, and the figures which resulted are presented in Table IV.

Explanation of Table IV is in order. The turnover of 132.25 is the average number of people who quit their jobs each week. This is a very large figure; however, ninety per cent of the total results from the undesirable tasks such as cleaning up. The figure for average days run per period, 4.5, is also totaled on a weekly basis. One should also note that the regression coefficients are relatively large for the two variables. Therefore, a regression line could be computed using the intercept and regression coefficients. The F value listed gives a significance of .02. It is very unlikely that the correlation coefficient and the coefficient of determination could have occurred by chance since the F value is so high.

The implications arising seem quite clear. There should be an attempt made to decrease the turnover as well as increase the days run per week in order to realize the maximum effect of the learning curve. Whether tangible, intangible, or both types of incentives should be given employees is not known; but either one or a combination of the two incentives could be given at some time in the future and the results measured.

TABLE IV

STATISTICS OF TURNOVER AND AVERAGE DAYS RUN PER PERIOD WHEN CORRELATED WITH THE DEPENDENT VARIABLE, COST

	Mean	Regression Coefficient	Multiple Correlation	Coefficient of Determination	F Value
Turnover	132.25	0.00030	.88	.77	9.23
Average Days Run Per Period	4.50	-0.00665			

Discussion of Breed of Turkey Processed, Mileage, Sanitation, and Indirect Labor

It was hoped in the initial stages of this study that some results could be found which would indicate whether or not different breeds of turkeys had varying influences on cost per pound. It was discovered though that data on only one grower's turkeys could be obtained. Consequently, the implications were not significant as compared to what had been hoped. Using the data obtained and expressing the number of turkeys received from the one grower as a percentage of total volume, a correlation coefficient of .36 occurred which, in turn, led to a coefficient of determination of only .11. Furthermore, the results were statistically significant at only the .35 level.

Mileage, or the distance which turkeys are hauled from the grower to the processor, is usually thought to be an important influence on cost. Reasons for this include shrinkage of turkeys, mortality, downgrades, as well as the obvious factor of truck costs. It was discovered that when mileage was placed against the dependent variable, a correlation of .55 and a coefficient of determination of .30 occurred. As a result, when taken individually, mileage accounted for .30 of the variance in cost. The suggestion from this study is that more production should be gathered from the state within which the newest poultry processing plant is located. However, it is recognized that in-state production was and still is very limited. The processor is forced then to go out of state to supplement in-state production. It seems that although mileage is important, it is not as crucial to the cost function as are other variables such as number of turkeys available, turnover, and the average days run per period.

Finally, two variables should be discussed which were thought to have very little influence on cost. These variables are sanitation and indirect plant expense.

Sanitation when correlated against cost showed a correlation of .77. The F value was 9.03 which proves significant at the .02 level. Indirect labor when correlated against cost gave a correlation of .81. Additionally, an F value of 11.52 was obtained which is significant at the .02 level.

When sanitation and indirect labor were multiplely correlated against cost, a correlation of .84 occurred. This led to a coefficient of determination of .70. The F value of 6.21 is significant at the .04 level.

The results from sanitation and indirect labor proved to be significant, rather than insignificant as the plant manager had thought would be the case. The exact reason for the results is not known; however, some suggestions may be given.

As sanitation and indirect labor expenses increase, working conditions in the plant might become more suitable.

Employees might then respond and turn out a greater output. This suggestion is not unreasonable. Studies have been conducted in the past which demonstrated that improved working conditions often lead to increased employee output.

Summary of Data Analysis

The major findings of this chapter include some rather surprising ones. For example, the independent variable of supervision had a correlation coefficient of only .01. This suggests (1) that processing line tasks, when once learned, become repetitive and that additional supervision will not increase plant efficiency, and (2) the task of supervision may be rather routine.

Sanitation showed high correlation in this study with regard to cost. As indicated, workers possibly may perceive more sanitary plant conditions and then perform their duties more efficiently. This would in turn force costs downward.

The concept of turnover and the associated learning curve are fairly new territories of thought. However, if constant labor losses occur, employees unfamiliar with their assigned job tasks will be always present. Since the poultry processing line is a type of assembly line, the production may proceed no more rapidly than the output of the least efficient worker. The suggestion that may be made is that turnover should be decreased. If this were accomplished, employees could operate in a more efficient manner as the learning curve effect materialized.

Implications for further research may be outlined. Foremost in importance, plant managers should keep more and better cost data so that they would be more informed on the factors which actually do influence cost significantly.

If at all possible, future studies should implement data obtained from several plants. Such data is difficult to obtain because of a lack of cooperation inherent among processing managers. However, a multiple plant study would allow an individual to determine the exact causes for differing plant efficiencies. This would allow suggestions to be made concerning the most feasible manner in which future plants should be constructed.

Foremost, a larger sample should be incorporated into future studies so that results will become more statistically significant. Increased significance allows one to predict with more confidence and lends creditability to the contents of a study.

CHAPTER V

SUMMARY AND CONCLUSIONS

Implications Resulting From the Study

As has been shown in the data analysis chapter, correlation and regression analysis are applicable to the poultry processing industry. Variables, which have been known to be of importance for some time, such as tonnage, may be examined in greater detail. In addition, and perhaps even more important, factors which have been thought to be of little importance as related to cost suddenly appear in a new light. Even within the relatively severe confinements which this study was conducted, the variables of sanitation and indirect labor were found to influence cost more than had been thought previously.

Furthermore, some variables which up to now have been held sacred come under scrutiny. For example, supervision in this study was found to be related very little to cost. It may well be the case that this was caused by the small sample size. However, it could also be that heavy supervision is not as critical as customarily thought. This is not to say that supervision is totally unimportant, for obviously some authority must be present for tasks to be accomplished. However, only limited supervision is apparently needed to insure that employees do not become too lax with respect to their assigned tasks.

Again, it must be emphasized that this study is intended to be exploratory in nature. Critics will contend that some limitations are present such as the small sample size, and indeed they will be correct. However, the important point is that one should visualize the potential uses of correlation and regression analysis in this industrial setting. The proper application and interpretation of such analysis can only lead to the ability to reduce internal costs of an existing plant. In addition, builders of future processing plants might well take note of the possibilities of these tools.

With the further refinement of regression and correlation analysis, many areas of possibility exist. With an increased sample size, one can take larger combinations of the independent variables and better determine the interrelationships existing within the cost function. Also, multiple plant studies may be made to explicitly determine why certain plants operate at lower costs than others.

This study has subjected certain cost-output relationships to analytical testing with the aid of regression and correlation analysis. There has always been statistical work in the sense of collecting and tabulating data on economic subjects; the newer emphasis, however, is on the application of statistical techniques to the data in order to estimate economic relationships and to test various hypotheses about such relationships.

It is hoped that in the future an increased body of empirically tested propositions concerning the cost-output relationship will be developed. There are two major difficulties which may hinder such development.

First, processing managers, as well as other representatives of the poultry processing industry, must come to realize that regression and correlation analysis can be a powerful tool with which to study in detail the cost function. Regression and correlation analysis is admittedly more difficult to understand initially than intuitive methods. However, once one is able to conceptualize the potential applications of such analysis, advice can easily be obtained from individuals learned in the field, such as university mathematicians, to determine in exactly what manner a given cost problem should be solved.

Secondly, when the time does come that processing managers agree to implement regression and correlation analysis, care should be taken that any answers determined are not held to be valid indefinitely. Cost relationships may change suddenly, especially if they depend on changing institutional arrangements or evolving patterns of human behavior. Thus, some part of each year's analytical work may have to be done anew.

This study has pointed out repeatedly the benefits of regression and correlation analysis. At the very least, individuals are forced to become more analytical as they implement this analysis. If only this one goal were to be accomplished, the use of regression and correlation techniques would be justified.

A SELECTED BIBLIOGRAPHY

Ezekiel, Mordecai. *Methods of Correlation Analysis*. London: John Wiley and Sons, Inc., 1941.

Ezekiel, Mordecai, and Karl A. Fox. *Methods of Correlation and Regression Analysis*. New York: John Wiley and Sons, Inc., 1959.

Fisher, Ronald A., and Frank Yates. *Statistical Tables for Biological, Agricultural, and Medical Research*. New York: Hafner Publishing Company, Inc., 1963.

Hoel, Paul G. *Elementary Statistics*. New York: John Wiley and Sons, Inc., 1966.

Johnston, John. *Statistical Cost Analysis*. New York: McGraw-Hill Book Company, Inc., 1960.

Lyle, Philip. *Regression Analysis of Production Costs and Factory Operations*. New York: Hafner Publishing Company, Inc., 1957.

Ostle, Bernard. *Statistics in Research*. Ames, Iowa: The Iowa State University Press, 1968.

Owen, D. B. *Handbook of Statistical Tables*. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1962.

APPENDIX A

QUESTIONNAIRE

1. What is the average cost per pound for each time period?
2. How many employees are utilized?
3. What is the turnover rate?
4. How many toms and hens have been processed?
5. What is the line speed for toms and hens?
6. What grower's birds are brought in to be processed?
7. How many downgrades occur and for what reasons?
8. How do the skill levels of employees differ and why?
9. Has any new equipment been purchased within the last year?
10. Have any new regulations affecting processing cost been introduced?
11. Have wage scales changed, and if so, by how much?
12. Have any additional supervisors been hired?
13. Have any labor disputes occurred?
14. What is the plant's total volume both in pounds and number of turkeys processed?
15. What is the average distance that turkeys are hauled from grower to processor?
16. How many days has the plant operated with the past year?

VITA

Roy C. Ferguson III

Candidate for the Degree of

Master of Business Administration

Thesis: AN APPLICATION OF LINEAR REGRESSION AND CORRELATION ANALYSIS TO THE POULTRY PROCESSING INDUSTRY

Major Field: Business Administration

Biographical:

Personal Data: Born in Shawnee, Oklahoma, October 5, 1946, the son of Mr. and Mrs. Roy C. Ferguson II.

Education: Graduated from Shawnee High School, Shawnee, Oklahoma, in May, 1964; received the Bachelor of Science degree from Oklahoma State University in 1968, with a major in General Business; completed requirements for the Master of Business Administration degree in May, 1970.

Professional Experience: Graduate Assistant in the College of Business Administration, Oklahoma State University, 1968-1969.

AN APPLICATION OF LINEAR REGRESSION AND CORRELATION ANALYSIS TO THE POULTRY PROCESSING INDUSTRY

Report Approved:

Kent Mingo
Report Adviser

George E. Pinches

Head, Department of Administrative Sciences

Name: Roy Coleman Ferguson III Date of Degree: May, 1970

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: AN APPLICATION OF LINEAR REGRESSION AND CORRELATION ANALYSIS TO THE POULTRY PROCESSING INDUSTRY

Pages in Study: 41 Candidate for Degree of Master of Business Administration

Major Field: Business Administration

Scope and Method of Study: Within the last decade, poultry processing plants have become much larger. Consequently, investment in such plants is increasing. However, most plant managers still do not possess an adequate, analytical knowledge of the cost function. Linear regression and correlation analysis is one method that may allow individuals to better understand the poultry processing cost function.

Findings and Conclusions: The proper use and application of linear regression and correlation analysis allows individuals to determine the exact influence which variables have on the cost function.

Volume has traditionally been thought to be the main determinant of cost. In this study, volume was indeed the primary factor influencing cost. Management, rather surprisingly, indicated very little correlation as related to cost. However, the variables of sanitation and indirect labor gave high correlations. Before this study was conducted, sanitation and indirect labor were thought to have an insignificant influence on costs.

A more detailed, analytical explanation of the poultry processing cost function is needed. With the aid of linear regression and correlation analysis, such explanation is now possible.

ADVISER'S APPROVAL Kent Mingo