CHAPTER 7

CORRELATION ANALYSIS

This chapter is concerned with regression or correlation analysis of the data obtained by interviewing 24 farmers. The purpose of the analysis is to lead to an understanding of the relationships between the explanatory variables and the dependent variable. The dependent variable (Y) was taken to be the trip attraction of a farm, that is the number of trips per year made by farmers visiting their farms. The seven explanatory or independent variables were chosen as the factors thought to affect the number of trips per year: distance from house to farm in km (X_1); household size as persons per household (X_2); farm area in rai (X_3); total farm income in Baht/1000 (X_4); income per rai in Baht per rai (X_5); number of 2-wheel vehicles owned (X_6); and number of 4-wheel vehicles owned (X_7). The hypothesis of this analysis is that the dependent varible, trips per year, is thought to be affected to some degree by each of the explanatory variables. The following analysis aims to prove or disprove this hypothesis.

It may be noted that the analysis classified vehicles owned into two types -- 2-wheel vehicles and $\geq$ 4-wheel vehicles -- as the dependent variables, X_6 and X_7. This is because of the different characteristics of these two vehicle types, their varying purposes, different operating costs, and the likelihood of their being used for trips of different lengths.

Method of Approach

Multiple linear regression analysis was used to find the correlation between the dependent variable and the independent variables. A computer program made available by the Department of Computers Engineering, Faculty of Engineering, Chulalongkorn University, was the means of solving this problem.

Data and Sample Size

Data from only 19 of the 24 interviewed households were utilized in the regression analysis. This is because there were 5 households that lived on their farms, so information on the trips per year to their farms was not pertinent. A listing of all variables is shown in Table 68.

Table 68 - Data for the Regression Analysis

Sample No.	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	300	8	8	20	7	350	4	0
2	200	8	8	18	22	1,220	1	0
3	50	15	6	10	20	2,000	2	1
4	180	15	5	830	1,875	2,260	1	10
5	240	4	6	57	20	350	1	0
6	240	8	11	410	720	1,760	2	3
7	300	1	6	40	40	1,000	3	0
8	180	22	4	150	260	1,730	2	2
9	100	15	6	27	113	4,180	0	2
10	300	15	10	30	17	570	2	0
11	100	15	7	55	55	1,000	2	0
12	60	8	4	8	8	1,000	3	0
15	250	2	8	30	29	970	0	0
17	180	5	5	160	332	2,080	1	3
18	360	8	7	450	660	1,470	1	3
19	300	15	7	350	660	1,880	2	3
20	50	8	3	150	282	1,880	1	2
21	340	4	5	65	85	1,310	1	0
24	150	22	5	150	65	430	2	2

Interviews deleted: Nos. 13, 14, 16, 22, 23

Where

Y = trips per year;

X_1 = distance from house to farm, km;

X_2 = household size, persons per household;

X_3 = farm area, rai;

X_4 = total farm income, Baht/1000;

X_5 = income per rai, Baht per rai;

X_6 = number of 2-wheel vehicles owned; and

X_7 = number of $\geq$ 4-wheel vehicles owned.

Formulation of the Equations

Three types of equations were formulated to estimate the number of trips per year made by the farmers from house to farm.

1. Linear form:

$$Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \text{-----}$$

2. Semi-logarithmic form:

$$\ln Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \text{------}$$

3. Exponential form:

$$Y = \beta_o X_1^{\beta_1} X_2^{\beta_2} X_3^{\beta_3} \text{-------}$$

Testing

All of the explanatory variables were individually compared to the dependent variable to find how they related to the dependent variable. Also, changes of the array of independent variables--by adding or withdrawing--were tested to find the equation of best fit. The accepted equations must have a high value of the correlation coefficient (R), a reasonable sign for the coefficient of each independent variable, and should satify the desired level of significance. Application of these criteria is called the "stepwise approach".

In the significance test of the coefficients of each independent variable, the Student's "t" test was used. The "t" distribution for degrees of freedom from 11 to 17, and levels of significance of 5 %, 10 %, and 20 %, are shown in Table 69.

Table 69 - Student's "t" Distribution

Df / LS.	11	12	13	14	15	16	17
5 %	2.201	2.179	2.160	2.145	2.131	2.120	2.110
10 %	1.796	1.782	1.771	1.761	1.753	1.756	1.740
20 %	1.363	1.356	1.350	1.345	1.341	1.337	1.333

Df. = Degrees of freedom

LS. = Level of significance

Source: R.A. Fisher, Statistical Methods for Research Workers, published by Oliver & Boyd, Ltd., Edinburgh.

Linear Form

The equations in Table 70 are the estimated values of Y affected by the different values of X. All 19 samples were used to formulate these equations. There are 26 linear equations that can be written to estimate the value of Y, but some of the equations derived were not reasonable in sign, or did not achieve the desired level of significance.

The best equation: (No. 9 in Table 70) is:

$$Y = 117.0 - 4.85 X_1 + 21.64 X_2 \qquad (I)$$

having $R^2 = 0.32$.

Equation (I) shows that the distance from house to farm (X_1) and household size (X_2) chiefly affect the number of trips per year. An increase of 1 km of distance would decrease the number of trips by 4.9 trips per year, at the 20 percent level of significance. If the household size were increased by one person, it would cause the number of trips to increase by 21.6 trips per year at the 20 percent level of significance. These two variables can explain 32 percent of the changes in the number of trips per year.

Equation (I) was obtained from the entire complement of 19 samples, but it may be seen that sample No. 4 in Table 68 is considerably different from the other samples. It is appropriate to delete this sample and repeat the linear-form analysis for 18 samples; this analysis follows.

Linear Form

Table 71 shows the estimated values of Y in the form of 17 equations. Sample No. 4 has been deleted for the present analysis. The stepwise approach was used to find

the equation of best fit. It was found that Equation No. 7 in Table 71 is the most effective equation.

$$Y = 114.7 - 5.58 X_1 + 18.78 X_2 + 0.22 X_3 \qquad \text{(II)}$$

having $R^2 = 0.41$

From Equation (II), it can be concluded that an increase of 1 km of distance from house to farm would reduce the dependent variable by 5.6 trips per year, at the 20 percent level of significance. An increase of one person in the household size would result in an increase of 18.8 trips per year. Also, an increase of one rai of farm area would cause trip attraction to increase by 0.22 trips per year. These three variables can explain 41 percent of the number of trips per year.

It should be noted that, after deleting the unusual sample, the new variable (X_3), had an effect on the dependent variable, and the R^2 value increased.

Semi-logarithmic Form

A total of 23 equations resulted from the analysis of the semi-logarithmic regression. Equation No. 8 in Table 72 has the best fit for estimating the number of trips per year.

$$\ln Y = 4.37 - 0.03 X_1 + 0.17 X_2 \qquad \text{(III)}$$

having $R^2 = 0.34$.

Taking the distance from house to farm with the household size, it is evident that an increase of 1 km of distance from house to farm would decrease the value of the logarithm of trips per year by 0.03 at the 20 percent level of significance. If the household size were to increase by one person, the value of ln (trips/year) would increase by 0.17. The combination of these two variables can explain 34 percent of the number of trips per year from house to farm.

Exponential Form

It was found that none of the equations shown in Table 73, which shows the regression equations in exponential form, were valid. None of these equations satisfied the stepwise tests. Equation No. 16 in Table 73 is seen to be nearly acceptable, but the negative sign of the exponent of variable X_5 is not reasonable; X_5 is the income per rai. This equation indicates that if the income per rai were higher, the trips per year must have fewer in number. Thus, there is no equation in Table 73 that satisfies the criteria of the statistical tests.

Table 70 - Equations in Linear Form (All 19 samples)

Eq. No.	Equation forms	R^2
1	$Y = 160.1 - \underset{(0.94)}{4.12}\ X_1 + \underset{(1.07)}{13.24}\ X_2 + \underset{(1.18)}{0.64}\ X_3 - \underset{(0.30)}{0.10}\ X_4 - \underset{(0.41)}{0.01}\ X_5 - \underset{(0.06)}{1.44}\ X_6 - \underset{(0.62)}{29.11}\ X_7$	0.15
2	$Y = 264.1 - \underset{(1.59)}{5.74}\ X_1$	0.13
3	$Y = 54.4 + \underset{(2.25)}{23.52}\ X_2^2$	0.23
4	$Y = 190.8 + \underset{(0.76)}{0.08}\ X_3$	0.03
5	$Y = 198.0 + \underset{(0.42)}{0.02}\ X_4$	0.01
6	$Y = 260.7 - \underset{(1.55)}{0.04}\ X_5$	0.12
7	$Y = 191.9 + \underset{(0.32)}{7.57}\ X_6$	0.006
8	$Y = 208.7 - \underset{(0.27)}{2.74}\ X_7$	0.004

Table 70 - Continued

Eq. No.	Equation forms	R^2
9	$Y = 117.0 - 4.85\ X_1 + 21.64\ X_2$ $(1.46)\quad(2.12)$	0.32
10	$Y = 109.5 - 5.62\ X_1 + 21.21\ X_2 + 0.11\ X_3$ $(1.68)\quad(2.11)\quad(1.18)$	0.38
11	$Y = 115.8 - 6.02\ X_1 + 17.99\ X_2 + 0.81\ X_3 - 0.33\ X_4$ $(1.88)\quad(1.83)\quad(1.78)\quad(1.56)$	0.47
12	$Y = 147.1 - 5.50\ X_1 + 17.08\ X_2 + 0.65\ X_3 - 0.25\ X_4 - 0.02\ X_5$ $(1.66)\quad(1.70)\quad(1.30)\quad(1.03)\quad(0.78)$	0.49
13	$Y = 143.9 - 5.57\ X_1 + 17.05\ X_2 + 0.66\ X_3 - 0.25\ X_4 - 0.02\ X_5 + 1.48\ X_6$ $(1.55)\quad(1.63)\quad(1.24)\quad(0.98)\quad(0.61)\quad(0.06)$	0.49
14	$Y = 41.9 + 23.43\ X_2 + 0.08\ X_3$ $(2.22)\quad(0.88)$	0.26
15	$Y = 169.6 + 0.90\ X_3 - 0.39\ X_4$ $(1.73)\quad(1.60)$	0.17
16	$Y = 259.7 + 0.06\ X_4 - 0.05\ X_5$ $(1.07)\quad(1.84)$	0.18
17	$Y = 291.4 - 0.05\ X_5 - 12.61\ X_6$ $(1.55)\quad(0.48)$	0.14

Table 70-Continued

Eq. No.	Equation forms	R^2
18	$Y = 197.2 + 6.38\ X_6 - 2.08\ X_7$ (0.25) (0.19)	0.008
19	$Y = 43.2 + 20.66\ X_2 + 0.71\ X_3 - 0.30\ X_4$ (1.96) (1.45) (1.31)	0.34
20	$Y = 96.8 + 19.04\ X_2 + 0.50\ X_3 - 0.18\ X_4 - 0.03\ X_5$ (1.80) (0.95) (0.73) (1.07)	0.39
21	$Y = 120.3 + 19.08\ X_2 + 0.47\ X_3 - 0.17\ X_4 - 0.03\ X_5 - 9.17\ X_6$ (1.75) (0.87) (0.65) (1.09) (0.37)	0.39
22	$Y = 160.5 + 11.26\ X_2 + 0.53\ X_3 + 0.05\ X_4 - 0.02\ X_5 - 9.46\ X_6 - 52.48\ X_7$ (0.93) (1.00) (0.17) (0.53) (0.39) (1.33)	0.47
23	$Y = 231.0 + 0.61\ X_3 + 0.10\ X_4 - 0.02\ X_5 - 9.33\ X_6 - 70.13\ X_7$ (1.16) (0.35) (0.46) (0.39) (2.05)	0.43
24	$Y = 275.0 + 0.39\ X_4 - 0.03\ X_5 - 12.89\ X_6 - 69.66\ X_7$ (2.25) (0.97) (0.54) (2.01)	0.37
25	$Y = 291.5 - 0.05\ X_5 - 12.37\ X_6 + 5.08\ X_7$ (1.57) (0.46) (0.45)	0.15
26	$Y = 156.5 - 4.21\ X_1 + 19.73\ X_2 - 0.02\ X_5$ (1.24) (1.90) (0.99)	0.36

Table 71 - Equations in Linear Form (18 Samples)

Eq. No.	Equation forms	R^2
1	$Y = 264.1 - 5.76\ X_1$ (1.52)	0.13
2	$Y = 52.9 + 23.68\ X_2$ (2.17)	0.23
3	$Y = 174.7 + 0.26\ X_3$ (1.50)	0.12
4	$Y = 183.1 + 0.12\ X_4$ (1.20)	0.08
5	$Y = 260.9 - 0.04\ X_5$ (1.49)	0.12
6	$Y = 114.1 - 5.03\ X_1 + 22.13\ X_2$ (1.45) (2.09)	0.32
7	$Y = 114.7 - 5.58\ X_1 + 18.78\ X_2 + 0.22\ X_3$ (1.65) (1.79) (1.42)	0.41
8	$Y = 114.4 - 6.16\ X_1 + 18.58\ X_2 + 0.93\ X_3 - 0.42\ X_4$ (1.85) (1.80) (1.60) (1.27)	0.47

Table 71 - Continued

Eq. No.	Equation forms	R^2
9	$Y = 153.8 - 5.33\ X_1 + 16.55\ X_2 + 0.55\ X_3 - 0.18\ X_4 - 0.02\ X_5$ (1.47) (1.51) (0.66) (0.35) (0.63)	0.49
10	$Y = 47.7 + 20.9\ X_2 + 0.19\ X_3$ (1.90) (1.18)	0.29
11	$Y = 169.0 - 0.82\ X_2 - 0.33\ X_4$ (1.22) (0.87)	0.17
12	$Y = 249.0 + 0.19\ X_4 - 0.06\ X_5$ (1.96) (2.16)	0.30
13	$Y = 42.0 + 20.94\ X_2 + 0.76\ X_3 - 0.34\ X_4$ (1.89) (1.21) (0.94)	0.34
14	$Y = 127.3 + 16.85\ X_2 + 0.13\ X_3 + 0.07\ X_4 - 0.04\ X_5$ (1.48) (0.16) (0.14) (1.20)	0.40
15	$Y = 168.9 - 4.90\ X_1 + 15.81\ X_2 + 0.26\ X_3 - 0.03\ X_5$ (1.49) (1.52) (1.72) (1.42)	0.49
16	$Y = 114.6 - 5.29\ X_1 + 19.77\ X_2 + 0.09\ X_4$ (1.53) (1.83) (1.02)	0.37
17	$Y = 155.6 - 4.43\ X_1 + 20.32\ X_2 - 0.03\ X_5$ (1.27 (1.90) (1.04)	0.37

Table 72 - Equations in Semi-logarithmic Form

Eq. No.	Equation forms	R^2
1	$\ln Y = 5.43 - \underset{(1.11)}{0.03 X_1}$	0.07
2	$\ln Y = 4.10 + \underset{(2.39)}{0.16 X_2}$	0.26
3	$\ln Y = 4.97 + \underset{(0.001)}{0.002 X_3}$	0.11
4	$\ln Y = 5.03 + \underset{(1.08)}{0.0007 X_4}$	0.07
5	$\ln Y = 5.51 - \underset{(1.45)}{0.0002 X_5}$	0.12
6	$\ln Y = 5.13 + \underset{(0.09)}{0.02 X_6}$	0.0005
7	$\ln Y = 5.16 - \underset{(0.04)}{0.006 X_7}$	0.0001
8	$\ln Y = 4.37 - \underset{(1.35)}{0.03 X_1} + \underset{(2.49)}{0.17 X_2}$	0.34

Table 72 - Continued

Eq. No.	Equation forms	R^2
9	$\ln Y = 4.35 - 0.03 X_1 + 0.15 X_2 + 0.001 X_3$ $(1.42) \quad (2.22) \quad (1.13)$	0.40
10	$\ln Y = 4.29 - 0.03 X_1 + 0.15 X_2 + 0.003 X_3 - 0.001 X_4$ $(1.06) \quad (2.15) \quad (0.73) \quad (0.48)$	0.41
11	$\ln Y = 4.63 - 0.02 X_1 + 0.13 X_2 + 0.0006 X_3 + 0.0005 X_4 - 0.0002 X_5$ $(0.86) \quad (1.77) \quad (0.10) \quad (0.15) \quad (0.81)$	0.44
12	$\ln Y = 4.87 - 0.02 X_1 + 0.13 X_2 - 0.00007 X_3 + 0.0009 X_4 - 0.0003 X_5 - 0.084 X_6$ $(0.64) \quad (1.65) \quad (0.01) \quad (0.25) \quad (0.91) \quad (0.47)$	0.45
13	$\ln Y = 4.07 + 0.15 X_2 + 0.001 X_3$ $(2.12) \quad (1.03)$	0.31
14	$\ln Y = 4.94 + 0.005 X_3 - 0.002 X_4$ $(1.21) \quad (0.89)$	0.15
15	$\ln Y = 5.43 + 0.001 X_4 - 0.0003 X_5$ $(1.77) \quad (2.03)$	0.27
16	$\ln Y = 5.79 - 0.0003 X_5 - 0.115 X_6$ $(1.56) \quad (0.66)$	0.14

Table 72-Continued

Eq. No.	Equation forms	R^2
17	$\ln Y = 5.14 + 0.014 X_6 - 0.003 X_7$ (0.08) (0.02)	0.0006
18	$\ln Y = 4.03 + 0.15 X_2 + 0.005 X_3 - 0.002 X_4$ (2.12) (1.22) (0.99)	0.36
19	$\ln Y = 4.50 + 0.13 X_2 + 0.001 X_3 - 0.00001 X_4 - 0.0002 X_5$ (1.72) (0.27) (0.004) (1.02)	0.40
20	$\ln Y = 4.86 + 0.11 X_2 + 0.002 X_3 + 0.005 X_4 - 0.0002 X_5 - 0.108 X_6 - 0.158 X_7$ (1.26) (0.27) (0.14) (0.67) (0.63) (0.51)	0.44
21	$\ln Y = 5.56 + 0.002 X_3 + 0.001 X_4 - 0.0002 X_5 - 0.111 X_6 - 0.31 X_7$ (0.30) (0.34) (0.65) (0.64) (1.05)	0.36
22	$\ln Y = 5.70 + 0.002 X_4 - 0.0003 X_5 - 0.128 X_6 - 0.268 X_7$ (1.86) (1.38) (0.79) (1.07)	0.36
23	$\ln Y = 5.79 - 0.0004 X_5 - 0.123 X_6 - 0.127 X_7$ (1.78) (0.70) (0.87)	0.19

Table 73 - Equations in Exponential Form

Eq. No.	Equation forms	R^2
1	$Y = 6.185\ X_1^{-0.430}$ (1.54)	0.13
2	$Y = 3.755\ X_2^{0.848}$ (1.18)	0.08
3	$Y = 4.465\ X_3^{0.201}$ (1.04)	0.06
4	$Y = 4.954\ X_4^{0.080}$ (0.50)	0.02
5	$Y = 9.009\ X_5^{-0.550}$ (4.96)	0.61
6	$Y = 4.713\ X_1^{-0.413}\ X_2^{0.791}$ (1.49) (1.14)	0.20
7	$Y = 3.825\ X_1^{-0.458}\ X_2^{0.783}\ X_3^{0.239}$ (1.68) (1.17) (1.32)	0.29

Table 73- Continued

Eq. No.	Equation forms	R^2
8	$Y = 3.890\ X_1^{-0.438}\ X_2^{0.693}\ X_3^{0.488}\ X_4^{-0.225}$ (1.56) (0.98) (1.12) (0.63)	0.31
9	$Y = 6.904\ X_1^{-0.226}\ X_2^{1.064}\ X_3^{-0.080}\ X_4^{0.327}\ X_5^{-0.613}$ (1.82) (3.45) (0.39) (1.92) (7.61)	0.88
10	$Y = 2.920\ X_2^{0.850}\ X_3^{0.201}$ (1.19) (1.06)	0.14
11	$Y = 4.348\ X_3^{0.608}\ X_4^{-0.367}$ (1.35) (1.01)	0.12
12	$Y = 8.478\ X_4^{0.227}\ X_5^{-0.615}$ (2.49) (6.16)	0.72
13	$Y = 2.920\ X_2^{0.850}\ X_3^{0.201}$ (1.19) (1.06)	0.14
14	$Y = 3.053\ X_2^{0.726}\ X_3^{0.520}\ X_4^{-0.287}$ (0.98) (1.14) (0.77)	0.18

Table 73 Continued

Eq.No.	Equation forms	R^2
15	$Y = 6.656 \underset{(3.29)}{1.100 X_2} \underset{(0.43)}{-0.094 X_3} \underset{(1.76)}{0.326 X_4} \underset{(7.58)}{-0.646 X_5}$	0.85
16	$Y = 6.660 \underset{(1.65)}{-0.225 X_1} \underset{(2.78)}{0.912 X_2} \underset{(3.16)}{0.277 X_3} \underset{(6.83)}{-0.548 X_5}$	0.84
17	$Y = 4.110 \underset{(1.60)}{-0.453 X_1} \underset{(1.21)}{0.847 X_2} \underset{(0.90)}{0.137 X_4}$	0.24
18	$Y = 7.602 \underset{(1.03)}{-0.180 X_1} \underset{(2.17)}{0.912 X_2} \underset{(5.19)}{-0.532 X_5}$	0.73
19	$Y = 4.348 \underset{(1.36)}{0.608 X_3} \underset{(1.01)}{-0.367 X_4}$	0.12
20	$Y = 8.478 \underset{(2.49)}{0.227 X_4} \underset{(6.16)}{-0.615 X_5}$	0.72