AGRICULTURAL ECONOMICS RESEARCH Vol. XIX, No. 4, OCTOBER 1967

Use of Marginal R^2 and Partial r^2 in a Multiple Regression Analysis

By Harry H. Harp

USE OF MULTIPLE REGRESSION techniques has been facilitated by the development of electronic computers and prewritten programs which make it possible to solve complex regression equations in minutes and at a greatly reduced cost. One of the problems remaining in the use of multiple regression is to find ways to present results in a way that is meaningful to the nonstatistician, as well as to those initiated in the use of statistical terms.

Presentation of the results of a multiple regression analysis is frequently more meaningful when the contribution of each independent variable in explaining the total R^2 is shown. This article reviews two techniques for calculating and presenting the contribution of each independent variable. Technique 1 is based upon the standardized regression coefficients (b*) adjusted for intercorrelation. Technique 2 is based upon the separate effect of individual variables. It is measured by observing the increment in explained variance when each variable is added after all other variables under consideration are entered and held constant.

Because of ease of computation, technique 1 has sometimes been used; but it is doubtful whether much confidence can be placed in it. Technique 2 has several advantages over technique 1. It is simpler to present in tabular form and it does not show negative contribution for variables when they actually contribute to explaining total variance. However, neither of the two techniques presented provides an unambiguous measure of the contribution of individual variables when interaction exists between the independent variables.

To illustrate the use of these techniques, data were taken from an analysis of the relationship between sales volume and factors influencing demand for 110 convenience foods.[1] The estimating equation derived from these data is as follows:

$$(1)\ \text{Log}\ \hat{Y} = -.60 - .60 (\log X_1)^2 - .85 \log X_2 + .28 (\log X_3)^2 + .31 \log X_4 + .65 \log X_5 - .16 (\log X_5)^2 + .44 \log X_6 + .23 \log X_7 - .58 \log X_8 + .33 \log X_9,$$

where the specific quantitative measures developed are as follows:

Sales:

$\hat{Y}$ = Estimated national sales of convenience foods in supermarkets in terms of 100 million servings sold annually.

Cost per serving:

X_1 = Cents per serving of convenience food.

Degree of competition:

X_2 = Sales of all other convenience items in same product group as percent of product group.

X_3 = Cents per serving of fresh or home-prepared foods.

X_4 = Cents per serving of highest volume competing convenience item in product group.

[1] Harry H. Harp and Marshall E. Miller, Convenience Foods: The Relationship Between Sales Volume and Factors Influencing Demand, U.S. Dept. Agr., Agr. Econ. Rpt. 81, Oct. 1965.

Importance in purchase pattern:

X_5 = National sales of all items in product group in supermarkets in terms of 100 million servings sold annually.

Availability:

X_6 = Percent availability of convenience items in terms of the percent of times observed by price enumerators in a sample of supermarkets in four metropolitan areas during a 12-month period of observation.

Success of similar convenience products:

X_7 = Sales of highest volume competing convenience item in same product group (100 million servings sold annually in supermarkets).

Special-product groups:

X_8 = Specialty products, i.e., foreign specialty products (1 if a specialty product, 0 if not).

X_9 = Ready-to-serve baked products (1 if a specialty product, 0 if not).

Other terms used in this paper include the following:

1. Simple r, the correlation between two variables with no restrictions on variables other than the two in question.
2. Partial r, one independent variable is correlated with the dependent variable and all other independent variables in the equation are statistically held constant at their mean.
3. Multiple R^2, the ratio of the variance explained by the regression equation to total variance.
4. R^2 delete, the multiple R^2 which would be obtained if a variable were deleted from the equation and the equation was recalculated.
5. Marginal R^2, the ratio of the increment in explained sum of squares contributed by an individual variable to the total sums of squares (or the multiple R^2 minus the R^2 delete) when the variable is added after all others under consideration are entered and statistically held constant at their mean.
6. Standardized coefficients (b^*), the b values transposed to standard units by multiplying the b values by the ratio of the standard deviation of the dependent and independent variable, $b^* = b\,(s_x/s_y)$.

Most of the terms and equations in this paper are known by statisticians and are often included in introductory textbooks on statistics. However, the term "marginal R^2" and the equations for computing this statistic are thought to be new.[2]

Technique 1

Technique 1 is based on standardized regression coefficients (b^*) adjusted for intercorrelation. The sum of these adjusted coefficients is equal to the multiple R^2. The direct and indirect effects of each independent variable on the multiple R^2 are computed from standardized coefficients (b^*'s) and simple r values.[3] Equation (2) shows that the direct effects of independent variables are the sum of the b^{*2} values, and the indirect effects are the sum of the products of 2 times the b^* values and the simple r's. If the independent variables have no intercorrelation among them, the indirect effects will be zero and the sum of the b^{*2} values will equal the multiple R^2. The equations for computing direct and indirect effects are illustrated with the market performance data previously identified.

$$
\begin{aligned}
(2)\ \text{Multiple } R^2 &= \text{direct effects} + \text{indirect effects} \\
&= (b_1^{*2} + b_2^{*2} + b_3^{*2} + b_4^{*2} + b_5^{*2} \\
&\quad + b_{52}^{*2} + b_6^{*2} + b_7^{*2} + b_8^{*2} + b_9^{*2}) \\
&\quad + (2b_1^* b_2 r_{12}) + (2b_1 b_3 r_{13}) \\
&\quad + (2b_1 b_4 r_{14}) + (2b_1^* b_5^* r_{15}) \\
&\quad + (2b_1^* b_{52}^* r_{152}) + (2b_1^* b_6^* r_{16}) \\
&\quad + (2b_1^* b_7^* r_{17}) + (2b_1^* b_8^* r_{18}) \\
&\quad + (2b_1^* b_9^* r_{19}) + (2b_2^* b_3^* r_{23}) \\
&\quad + (2b_2^* b_4^* r_{24}) + (2b_2^* b_5^* r_{25}) \ldots \\
&\quad + (2b_8^* b_9^* r_{89})
\end{aligned}
$$

---

[2] Equation (10), p. 109.

[3] Robert Ferber, Statistical Techniques in Marketing Research, McGraw-Hill, New York, 1st ed., Apr. 1949, p. 364.

Importance in purchase pattern:

X_5 = National sales of all items in product group in supermarkets in terms of 100 million servings sold annually.

Availability:

X_6 = Percent availability of convenience items in terms of the percent of times observed by price enumerators in a sample of supermarkets in four metropolitan areas during a 12-month period of observation.

Success of similar convenience products:

X_7 = Sales of highest volume competing convenience item in same product group (100 million servings sold annually in supermarkets).

Special-product groups:

X_8 = Specialty products, i.e., foreign specialty products (1 if a specialty product, 0 if not).

X_9 = Ready-to-serve baked products (1 if a specialty product, 0 if not).

Other terms used in this paper include the following:

1. Simple r, the correlation between two variables with no restrictions on variables other than the two in question.
2. Partial r, one independent variable is correlated with the dependent variable and all other independent variables in the equation are statistically held constant at their mean.
3. Multiple. R^2, the ratio of the variance explained by the regression equation to total variance.
4. R^2 delete, the multiple R^2 which would be obtained if a variable were deleted from the equation and the equation was recalculated.
5. Marginal R^2, the ratio of the increment in explained sum of squares contributed by an individual variable to the total sums of squares (or the multiple R^2 minus the R^2 delete) when the variable is added after all others under consideration are entered and statistically held constant at their mean.
6. Standardized coefficients (b^*), the b values transposed to standard units by multiplying the b values by the ratio of the standard deviation of the dependent and independent variable, $b^* = b\,(s_x/s_y)$.

Most of the terms and equations in this paper are known by statisticians and are often included in introductory textbooks on statistics. However, the term "marginal R^2" and the equations for computing this statistic are thought to be new.[2]

Technique 1

Technique 1 is based on standardized regression coefficients (b^*) adjusted for intercorrelation. The sum of these adjusted coefficients is equal to the multiple R^2. The direct and indirect effects of each independent variable on the multiple R^2 are computed from standardized coefficients (b^*'s) and simple r values.[3] Equation (2) shows that the direct effects of independent variables are the sum of the b^{*2} values, and the indirect effects are the sum of the products of 2 times the b^* values and the simple r's. If the independent variables have no intercorrelation among them, the indirect effects will be zero and the sum of the b^{*2} values will equal the multiple R^2. The equations for computing direct and indirect effects are illustrated with the market performance data previously identified.

$$
\begin{aligned}
(2)\ \text{Multiple } R^2 &= \text{direct effects} + \text{indirect effects} \\
&= (b_1^{*2} + b_2^{*2} + b_3^{*2} + b_4^{*2} + b_5^{*2} \\
&\quad + b_{5_2}^{*2} + b_6^{*2} + b_7^{*2} + b_8^{*2} + b_9^{*2}) \\
&\quad + (2b_1^* b_2 r_{12}) + (2b_1 b_3 r_{13}) \\
&\quad + (2b_1 b_4 r_{14}) + (2b_1^* b_5^* r_{15}) \\
&\quad + (2b_1^* b_{5_2}^* r_{15_2}) + (2b_1^* b_6^* r_{16}) \\
&\quad + (2b_1^* b_7^* r_{17}) + (2b_1^* b_8^* r_{18}) \\
&\quad + (2b_1^* b_9^* r_{19}) + (2b_2^* b_3^* r_{23}) \\
&\quad + (2b_2^* b_4^* r_{24}) + (2b_2^* b_5^* r_{25}) \ldots \\
&\quad + (2b_8^* b_9^* r_{89})
\end{aligned}
$$

---

[2] Equation (10), p. 109.

[3] Robert Ferber, Statistical Techniques in Marketing Research, McGraw-Hill, New York, 1st ed., Apr. 1949, p. 364.

Table 1 shows net effects as the difference between the direct and the indirect effects. Equation (3) may be used as a check of the net effect presented in table 1.

When a variable is added to the regression, all standardized coefficients (b*) are likely to change and true net effect of such an addition will be positive. However, difficulties have been encountered in using table 1 to show the direct, indirect, and net effect of variables in a regression equation. When there is a great deal of interaction between the independent variables, the so-called net effect shown in table 1 might appear to be negative if one were to forget that the direct, indirect, and net effects are integral parts of the sums and consider only a portion of these sums. The net effects for independent variables in a multiple regression analysis will appear to be negative when the signs of the b* and corresponding simple r are not the same (equation 3).

$$(3)\quad R^2 = \text{net effect } X_1 + \text{net effect } X_2 + \text{net effect } X_3 + \text{net effect } X_4 + \text{net effect } X_5 + \text{net effect } X_5^2 + \text{net effect } X_6 + \text{net effect } X_7 + \text{net effect } X_8 + \text{net effect } X_9$$

$$= b_1^* r_{y1} + b_2^* r_{y2} + b_3^* r_{y3} + b_4^* r_{y4} + b_5^* r_{y5} + b_{5^2}^* r_{y5^2} + b_6^* r_{y6} + b_7^* r_{y7} + b_8^* r_{y8} + b_9^* r_{y9}$$

In a multiple regression analysis the coefficients (b values) usually have the same sign as the simple r with the dependent variable. However, when two or more of the independent variables are highly correlated, the signs of the b values of the dependent variables may not agree with the signs for the corresponding simple r values.[4] This results because the weaker independent variables modify the effect of the stronger independent variable on the dependent variable. For instance, even though all of the variables in equation (1) made a statistically significant contribution to explaining variance in sales of convenience foods, three of the variables, X_2, X_3, and X_5^2, appear to have negative net effects in table 1. This is due to the intercorrelation among the independent variables.

When a negative sign is obtained for a net effect, it does not mean that the variable adds less than nothing to the reduction of unexplained variance or is of no significance. It means that the influence of the variable is working counter to the influence of other variables to reduce the bias in the final estimate. Thus, it may prevent some of the predicted values from going as low as they otherwise would when the effect of one or more of the other variables is downward, and it may tend to keep predicted values from going as high as they otherwise would when the other variables are forcing predicted values up.

Technique 1 is quite similar to the technique of separate determination which is described by Ezekiel.[5] The coefficient of separate determination is identical to the so-called net effect. The main difference between separate determination and technique 1 is that technique 1 attempts to allocate the influence of each variable into direct and indirect effects; whereas separate determination does not distinguish between the direct and indirect effects.

In an economic analysis, there is generally considerable interaction between some of the independent variables.[6] When such interaction exists, technique 1 will provide answers which lack clarity and are difficult to present in a nontechnical tabular form. Therefore, other measures of the individual importance of the independent variable--for example, technique 2--may be preferred.

---

[4] Karl A. Fox and James F. Cooney, Jr., Effects of Intercorrelation Upon Multiple Correlation and Regression Measures, U.S. Dept. Agr., AMS-341, 28 pp., 1954.

[5] Mordecai Ezekiel, Methods of Correlation Analysis, John Wiley and Sons, New York, 2d ed., 1941, p. 498.

[6] James N. Morgan and John A. Sonquist, "Problems in the Analysis of Survey Data and a Proposal," Jour. Amer. Statis. Assoc., Vol. 58, June 1963, pp. 415-435. Donald E. Farror and Robert R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited." Rev. Econ. and Statis., Vol. XLIX, No. 1, Feb. 1967, pp. 92-107.

279-249 O-67—2

Table 1.--Direct and indirect effects of the factors influencing sales of convenience foods

Effect	Cost per serving X_1	Convenience items as percent of product group X_2	Cost of home-prepared foods X_3	Cost of competing convenience foods X_4	Sales of product group X_5	Squared value of product group X_5^2	Availability of convenience foods X_6	Sales of competing products X_7	Specialty products X_8	Ready-to-serve baked products X_9	Total
Direct.....	.387	1.019	.069	.408	.340	.075	.017	.188	.054	.026	2.583
Indirect:											
X_1 and X_2	-.025	-.025									-.050
X_1 and X_3	-.141		-.141								-.282
X_1 and X_4	-.054			-.054							-.108
X_1 and X_5	.186				.186						.372
X_1 and X_5^2	-.062					-.062					-.124
X_1 and X_6	.018						.018				.036
X_1 and X_7	.043							.043			.086
X_1 and X_8	.068								.068		.136
X_1 and X_9	.009									.009	.018
X_2 and X_3		.019	.019								.038
X_2 and X_4		-.557		-.557							-1.114
X_2 and X_5		-.127			-.127						-.254
X_2 and X_5^2		.020				.020					.040
X_2 and X_6		.027					.027				.054
X_2 and X_7		-.398						-.398			-.796
X_2 and X_8		.012							.012		.024
X_2 and X_9		-.033								-.033	-.066
X_3 and X_4			.029	.029							.058
X_3 and X_5			-.077		-.077						-.154
X_3 and X_5^2			.023			.023					.046
X_3 and X_6			-.006				-.006				-.012
X_3 and X_7			-.019					-.019			-.038
X_3 and X_8			-.032						-.032		-.064
X_3 and X_9			-.009							-.009	-.018

X_4 and X_5				.058	.058						.116
X_4 and X_5^2				-.006		-.006					-.012
X_4 and X_6				-.010			-.010				-.020
X_4 and X_7				.230				.230			.460
X_4 and X_8				-.040					-.040		-.080
X_4 and X_9				.014						.014	.028
X_5 and X_5^2					-.137	-.137					-.274
X_5 and X_6					.009		.009				.018
X_5 and X_7					.117			.117			.234
X_5 and X_8					.031				.031		.062
X_5 and X_9					.009					.009	.018
X_5^2 and X_6						-.005	-.005				-.010
X_5^2 and X_7						-.043		-.043			-.086
X_5^2 and X_8						-.012			-.012		-.024
X_5^2 and X_9						-.001				-.001	-.002
X_6 and X_7							-.009	-.009			-.018
X_6 and X_8							.005		.005		.010
X_6 and X_9							-.003			-.003	-.006
X_7 and X_8								.004	.004		.008
X_7 and X_9								.014		.014	.028
X_8 and X_9									.003	.003	.006
Total indirect..	.042	-1.062	-.213	-.336	.069	-.223	.026	-.061	.039	.003	-1.716
Net........	.429	-.043	-1.44	+072	.409	-.148	.043	.127	.093	.029	.867

Technique 2

Technique 2 provides a measure of the additional variance explained when the variable is added to the regression equation after all of the other independent variables have been entered into the equation. Conceptually, the additional contribution of each independent variable after all others are included in the equation appears to be one of the most useful methods of presenting the importance of each variable in a regression equation. One such measure is the partial r^2 which shows the percentage each variable reduces the total unexplained variance after all other variables under consideration were previously entered and held constant.

Ezekiel and Fox have demonstrated that the partial r values are a measure of the separate effects of individual variables.[7] Equation (4) has been used to show this relationship. This technique corresponds to the last step in the stepwise regression:

$$(4)\quad r^2_{12.34} = 1 - \frac{1 - R^2_{1.234}}{1 - R^2_{1.34}}$$

$$r^2_{13.24} = 1 - \frac{1 - R^2_{1.234}}{1 - R^2_{1.24}}$$

$$r^2_{14.23} = 1 - \frac{1 - R^2_{1.234}}{1 - R^2_{1.23}}$$

Kenneth J. McCallister, Marketing Economics Division, ERS, shows the partial r^2 may be expressed as a function of the F value[8] for that particular variable and degrees of freedom involved. This significant equation is as follows:

$$(5)\quad \text{Partial } r^2 = \frac{1}{1 + \frac{\text{degrees of freedom}}{F}}$$

[7] Mordecai Ezekiel and Karl A. Fox, Methods of Correlation and Regression Analysis, John Wiley and Sons, New York, 3d ed., 1965, p. 192.

[8] $F = b^2/Sb^2$.

Equation (6), the extension of McCallister's equation for partial r^2, shows its relationship to Waugh's[9] equation for partial r. To simplify the notation, the partial regression coefficient is designated b, its standard error as Sb, degrees of freedom as d.f.

$$(6)\quad \text{Partial } r^2 = \frac{1}{1 + \frac{d.f.}{F}}$$

$$= \frac{F}{F + d.f.}$$

$$= \frac{\frac{b^2}{Sb^2}}{\frac{b^2}{Sb^2} + d.f.}$$

$$= \frac{b^2}{b^2 + (d.f.)(Sb^2)}$$

$$\text{Partial } r = \frac{b}{\sqrt{b^2 + (d.f.)(Sb^2)}}$$

Although partial r's and F values with fixed degrees of freedom are quite similar, some statisticians are accustomed to analyzing data and thinking in terms of one or the other. Thus, it may be desirable to present both partial r^2 and the F values or t values.[10]

From the partial r^2 and the multiple R^2, the R^2 delete may be computed. As previously defined, the R^2 delete is the multiple R^2 which would be obtained if a variable were deleted from the equation and the equation recalculated. The R^2 delete for each variable may be computed from the partial r^2 and the multiple R^2 as in the following equation:[11]

$$(7)\quad R^2 \text{ delete} = \frac{\text{multiple } R^2 - \text{partial } r^2}{1 - \text{partial } r^2}$$

[9] Frederick V. Waugh, "The Computation of Partial Correlation Coefficients," Jour. Amer. Statis. Assoc., Vol. 41, No. 236, Dec. 1946, pp. 543-546.

[10] The square of the 2-tailed t value at a given probability level is equal to the single-tailed F value.

[11] Mich. State Univ. Agr. Expt. Sta., Calculation of Least Squares (Regression) Problems on the LS Routine, STAT Ser. Descr. 7, Dec. 1966, p. 38.

The difference between the multiple R^2 and R^2 delete is the marginal R^2 for the deleted variable. The formula for computing the marginal R^2 is:

(8) Marginal R^2 = Multiple R^2 - R^2 delete

An alternate equation is:

(9) $\text{Marginal } R^2 = F \frac{\text{Residual mean square}}{\text{Total sum of squares}}$

This reduces to:

(10) $\text{Marginal } R^2 = \frac{\text{Reduction in sum of squares}}{\text{Total sum of squares}}$

The marginal R^2 values in tables 2 and 3 measure the increment in explained variance when each variable is added after all others under consideration are entered and statistically held constant at their mean.

Because the marginal R^2 values are all expressed as a ratio to the same base--i.e., the total sum of squares--they are directly comparable to each other and are additive. This permits summing individual marginal R^2 and subtracting from the total multiple R^2 to give a measure of the joint effects. This is shown in table 3.

(11) Joint effects = R^2 - sum of marginal R^2 values

In conclusion, the additional contribution of each variable in explaining total variance after all others are included in the equation appears to offer the least complex method of presenting the influence of each variable.

Table 2.--Relative importance of individual variables affecting sales of convenience foods

$R_{y123455^2 6789}$	Variable deleted	Variables in equation	R^2 delete	Marginal R^2 [1]
(1)	(2)	(3)	(4)	(5)
.87	X_2	$X_1, X_3, X_4, X_5, X_5^2, X_6, X_7, X_8, X_9$	.78	.09
.87	X_1	$X_2, X_3, X_4, X_5, X_5^2, X_6, X_7, X_8, X_9$	.78	.09
.87	X_5	$X_1, X_2, X_3, X_4, X_5^2, X_6, X_7, X_8, X_9$	.81	.06
.87	X_4	$X_1, X_2, X_3, X_5, X_5^2, X_6, X_7, X_8, X_9$	.82	.05
.87	X_8	$X_1, X_2, X_3, X_4, X_5, X_5^2, X_6, X_7, X_9$	.84	.03
.87	X_9	$X_1, X_2, X_3, X_4, X_5, X_5^2, X_6, X_7, X_8$	.85	.02
.87	X_5^2	$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$	.86	.01
.87	X_6	$X_1, X_2, X_3, X_4, X_5, X_5^2, X_7, X_8, X_9$	.86	.01
.87	X_3	$X_1, X_2, X_4, X_5, X_5^2, X_6, X_7, X_8, X_9$	.86	.01
.87	X_7	$X_1, X_2, X_3, X_4, X_5, X_5^2, X_6, X_8, X_9$	.86	.01

[1] Col. 5 = Col. 1 - Col. 4.

Table 3.--Contribution of each variable to explaining variance in sales of convenience foods

Source of variation	Sum of squares	F Value	Partial r^2 (1)	Marginal R^2
Independent effects:				
Convenience items as a percent of product group, X_2	3.76	66.79	.40	.090
Cost per serving, X_1	3.70	65.68	.40	.088
Product group, X_5	2.49	44.18	.31	.060
Cost of competing convenience items, X_4	2.13	37.92	.28	.051
Specialty products, X_8	1.34	23.78	.19	.032
Ready-to-serve baked products, X_9	.93	16.53	.14	.022
Square value of product group, X_5^2	.57	10.09	.09	.013
Availability, X_6	.55	9.79	.09	.013
Cost of home-prepared foods, X_3	.53	9.50	.09	.013
Sales of competing products, X_7	.50	8.82	.09	.012
Total independent effects	16.50			.394
Joint effects	19.78			.473
Total effects explained	36.28			.867
Residual variance	5.57			.133
Total	41.85			1.00

[1] With 99 degrees of freedom.