# Squared Versus Unsquared Deviations for Lines of Best Fit

By Harold B. Jones and Jack C. Thompson[2]

REGRESSION AND CORRELATION are widely used and commonly accepted as a basis for work in many applied fields. These techniques are usually based on the principle of least squares. The method of least squares, however, involves minimum squared deviations, and is subject to a number of inherent characteristics that differ from those of minimum unsquared deviations. The differences in the two concepts are frequently unrecognized or ignored except in studies oriented primarily toward mathematical theory (1).[3] The purpose of this paper is to compare and contrast the two approaches in the hope that more effective utilization of both techniques will result.

Standard textbooks often state that the least squares method provides the line of best fit, and imply that there is only one line of best fit for a given set of data. For example, "one must choose that line which 'best' fits the data . . . Our criterion of 'best' is the least-squares criterion" (11, p. 163). Many researchers and students have accepted least squares as a working tool without further questioning. It is the apparent widespread acceptance of this method as the only reliable means for establishing the true relationship between variables that has prompted this paper. In reality, the least squares relationship is only one of a number of possible relationships, each of which has its own assumptions and biases.

## The Central Problem

The method of least squares originated from mathematical theories developed by astronomers in the early 1800's for the purpose of determining the paths of comets and planets. These theories were an outgrowth of early probability theory suggested by Laplace and later modified by Legendre and Gauss (14, pp. 92-95). The early theories were combined with the later work of Galton on regression analysis (1889) to form the basic foundation upon which modern correlation and regression techniques rest.

From a mathematical standpoint the least squares method rests on one rather fundamental point: "that a number w will be called the best approximation to a set of numbers $(x_1, x_2 \ldots x_n)$, or the best representative for the set, in case the sum of the squares of the deviations of the x's from w is less than the sum resulting if w is replaced by any other number" (6, p. 330). Furthermore, "in view of the possibility of other definitions of a best approximation, we shall say that Definition I describes the best approximation in the sense of least squares." Thus these basic definitions point out two critical assumptions that underlie the principle of least squares: (1) that it is a method of approximation, and (2) that it is the best only in the sense of least squares. If we want to measure deviations in terms of actual data or cubed data or logarithms

[1] Submitted as Journal Paper No. 26, University of Georgia College of Agriculture Experiment Stations, College Station, Athens.

[2] This paper represents a joint contribution of the authors with no attempt to establish senior authorship. Ideas expressed do not necessarily imply endorsement by the University of Georgia or the U.S. Department of Agriculture.

[3] Underscored numbers in parentheses refer to items in the Literature Cited, p. 69.

Editor's note: As working economists we need to remind ourselves now and then that the choice of an appropriate line of best fit may depend more on the characteristics of the relationships we are measuring than on the statistical techniques with which we may be most familiar. This paper is intended to help the general but less statistically minded economist better understand a problem that may already be clear to the statistical specialist, and thus to choose more efficient working methods.

rather than squared data, then the best approximation may be entirely different. It is these two points which are crucial to a clear understanding of least squares analysis in relation to any alternative method.

One of the major advantages of using the least squares method is that it will provide the most probable estimate of the underlying relationship between certain factors when all other variables, including errors of measurements, are omitted. In other words, the method has predictive power, at least in a probability sense. The question is--how do you interpret what is the most probable estimate? Historically observed facts are one thing but future changes are another. Statistical inference and probability theory are highly interrelated. Yet the attempt to substitute probability for logic or cause-and-effect relationships carries one beyond the realm of true scientific inquiry.[4] This line of reasoning is more fully explained by Waugh, who states that "unless one has faith in the crystal ball or the Ouija board, he can never know what would have been true if some forces had been different. We are therefore forced to guess what would have happened" (15, p. 307). He goes on to state that "students more often put too much faith in the results of least squares than too little. They think that somehow the mathematical processes of the least-squares method give them an answer that is 'correct,' rather than an estimate or guess of what is correct."

Ezekiel and Fox recognize that "the least-squares line gives the line of best fit under the assumptions of that method: a normal distribution of the observations around the line and the reduction of the squared residuals to a minimum" (3, p. 68). However, it has been shown by the Markoff theorem that the assumption of normality is not necessarily essential to the theory of least squares (2, p. 105). But there does have to be a distribution of some kind which is based on the existence of a random variable (y) and which is independent of any of the other variables considered (x's). The least squares assumption thus becomes the relevant criterion when these conditions are met.

Another theoretical advantage of least squares is that the method is mathematically rigorous and thereby reduces the errors of measurement when compared with more subjective measures. In other words, it is a more consistent method of estimating. Yet, it does not necessarily follow that a consistent estimate is more accurate in describing a given relationship than an inconsistent estimate. Subjective methods of measurement may be more accurate even though less consistent than other methods. This reflects the old conflict of "precision" versus "accuracy." Is it better to be "approximately right" or "precisely wrong"? This point is well stated by A. N. Whitehead, the noted philosopher: "There is no more common error than to assume that, because prolonged and accurate mathematical calculations have been made, the application of the result to some fact of nature is absolutely certain" (9, p. 271).

Regardless of the assumptions involved, most statistical authorities have emphasized the usefulness of least squares in measuring the deviation of items about a mean or a line of best fit. Snedecor states that the simple average of individual variations is not relevant because it leads into a blind alley so far as statistical theory is concerned (10, pp. 36-37). Yet when considering why the deviations should be squared he says that, "in a non-mathematical discussion, it is quite impossible to give an adequate answer to this question."

Thus, the question of squaring deviations has usually been considered to hinge upon advanced statistical theory; perhaps not enough thought has been given to the judgment or logic to be used in individual situations that may not require advanced statistical technique.

## A Hypothetical Example

The following hypothetical example was designed to illustrate the differences in results obtained when the best approximation in terms of least squares is compared with the best approximation in terms of least absolute deviations.

The example represents a simplified case in which a relationship exists between X and Y and the objective is to predict values of Y from

[4] The validity of the inductive approach is at best based on highly problematical grounds and has been the subject of philosophical controversy for many centuries. See Hume's essay (8) first published in 1777. Fisher calls this inverse probability and states that "the theory of inverse probability is founded upon an error, and must be wholly rejected" (4, p. 9).

the values of X. Fitting a line on the basis of least squares gives the type of relationship shown in figure 1a. Fitting a line on the basis of least actual deviations by a freehand or judgment method gives the relationship in figure 1b. The basic data for these charts are given in table 1. This table shows that the least squares approach provides a line where the sum of the deviations without regard to sign is nearly twice the sum of the unsquared deviations. The sum of the deviations from the line of regression fitted by the least squares technique is 18.5 points whereas the sum of the deviations from the line based on the unsquared method is 10 points.

Carried one step further with both sets of estimates evaluated on the basis of squared deviations, the least squares technique gives the lower total sum of squares with a correspondingly lower average (column 7, table 1). Transposed into the traditional measure of correlation this provides a coefficient of determination ($r^2$) of 0.75 for the least squares method and an $r^2$ of 0.62 for the unsquared method (table 2). Only 62 percent of the actual variation is explained by this line whereas the coefficient of determination indicates that 75 percent of the squared variation (variance) is explained. However, if the

**TWO METHODS OF FITTING A REGRESSION LINE**

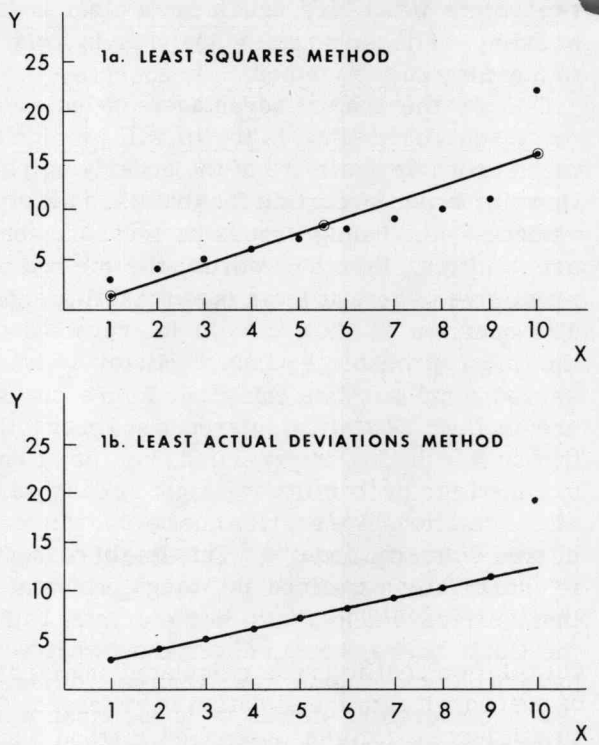


Figure 1

Table 1.--Basic data for calculation of regression equations and correlation coefficients by squared and unsquared methods

| Basic data | | | | Least squares method[a] | | | | | Non-squared method | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| X | Y | XY | $X^2$ | Yc | Y-Yc | $(Y-Yc)^2$ | $Y-\overline{Y}$ | $(Y-\overline{Y})^2$ | Yc | Y-Yc | $(Y-Yc)^2$ |
| 1 | 3 | 3 | 1 | 1.5 | 1.45 | 2.10 | 5.5 | 30.25 | 3 | 0 | 0 |
| 2 | 4 | 8 | 4 | 3.1 | .91 | .83 | 4.5 | 20.25 | 4 | 0 | 0 |
| 3 | 5 | 15 | 9 | 4.6 | .36 | .13 | 3.5 | 12.25 | 5 | 0 | 0 |
| 4 | 6 | 24 | 16 | 6.2 | .18 | .03 | 2.5 | 6.25 | 6 | 0 | 0 |
| 5 | 7 | 35 | 25 | 7.7 | .73 | .53 | 1.5 | 2.25 | 7 | 0 | 0 |
| 6 | 8 | 48 | 36 | 9.3 | 1.27 | 1.61 | .5 | .25 | 8 | 0 | 0 |
| 7 | 9 | 63 | 49 | 10.8 | 1.82 | 3.31 | .5 | .25 | 9 | 0 | 0 |
| 8 | 10 | 80 | 64 | 12.4 | 2.36 | 5.57 | 1.5 | 2.25 | 10 | 0 | 0 |
| 9 | 11 | 99 | 81 | 13.9 | 2.91 | 8.47 | 2.5 | 6.25 | 11 | 0 | 0 |
| 10 | 22 | 220 | 100 | 15.5 | 6.55 | 42.90 | 13.5 | 182.25 | 12 | 10 | 100 |
| Σ 55 | 85 | 595 | 385 | -- | 18.54 | 65.48 | 36.0 | 262.50 | -- | 10 | 100 |
| M 5.5 | 8.5 | -- | -- | -- | -- | 6.55 | -- | 26.25 | -- | -- | 10.00 |

[a] Deviations expressed without regard to signs. Certain columns rounded.

Table 2.--Coefficients of correlation and determination based on squared and unsquared deviations

| Type of regression line | Value of coefficients | |
|---|---|---|
| | $r^2$ | c |
| Least squared deviations... | 0.75 | 0.49 |
| Least actual deviations.... | .62 | .72 |

Coefficients based on the following formulas:

$$r^2 = \frac{\sigma y^2 - Sy^2}{\sigma y^2}$$

$$c = \frac{\Sigma(1Y-\overline{Y}1) - \Sigma(1Y-Yc1)}{\Sigma(1Y-\overline{Y}1)}$$

where $r^2$ = coefficient of determination and c = correlation coefficient based on unsquared deviations.

correlation coefficient is calculated on the basis of unsquared data the situation is reversed. The coefficient c for the unsquared method would then be 0.72 and the c for the least squares line would be 0.49. In this case then, 72 percent of the actual variation is explained and only 49 percent of the squared variation.

Obviously the least squares method provides the line of best fit when best fit is interpreted in terms of least squares. This is a circular process that defines the line of best fit in terms of one criterion and then evaluates the effectiveness of the fit in terms of the same criterion. This procedure yields an optimum line of regression when least squares are the appropriate criteria. By shifting the line on a trial and error basis, it is frequently possible to improve the accuracy of the actual predictions of Y from given values of X, but this would not be logical unless justified by the underlying relationships.

Even if the least squares criteria are accepted, there is still the problem of selecting the type of line which best represents the data being analyzed. The real significance of the correlation coefficient will depend not only on the goodness of fit, but on the type of relationship that is presumed to exist. A priori knowledge becomes extremely important here. Otherwise, one could not know whether the data are best represented by a straight line relationship, a curvilinear relationship, a relationship linear in the logarithms, or one of many other types of relationships that could exist between variables. If the nature of the relationship is not known and the wrong type of curve is fitted, the explanatory value will be relatively poor. This could still be the "line of best fit" as determined by the statistical method selected, but this would be no indication of the true underlying relationship, it would only mean that you have the best fitting line for that particular type of curve.

## Another Approach

If the primary objective is to predict values of Y from values of X in terms of minimum actual deviations rather than minimum squared deviations, other methods than that of least squares may be appropriate. However, in certain special cases the results may be the same. Where the distribution of errors is such that there is a counterbalancing effect on either side of the line of regression, then minimizing squared deviations will result in minimum actual deviations (note that the errors could be, but do not necessarily have to be, in the form of a normal distribution). In too many cases, however, minimum squared deviations are used when the evidence does not suggest the presence of a "balanced" or normal distribution.

In such situations, it may be better to try to minimize actual unsquared deviations by an iterative process similar to that already described, either by starting with a least squares solution or a group average method and working toward an optimum solution by the graphic method, or by more advanced linear programming techniques (see 7, p. 239, and 13). The regression coefficients could be calculated from the indicated functional relationships, and a correlation coefficient could be computed in terms of c rather than r where c is defined as:

$$c = \frac{\text{average deviation from mean} - \text{average deviation from regression line}}{\text{average deviation from mean}}$$

based on unsquared deviations (see footnote to table 2 for a statement of this formula in more familiar terminology).

These methods make it necessary to disregard signs, but they do provide a workable solution which could have a considerable advantage over the traditional method. They also allow the possibility of using the median rather than the mean as the base point from which to measure deviations. Since the median is the middle point, it has the useful property of being that point around which the sum of the absolute deviations is minimized.[5] Although the median is not as stable as the mean from a mathematical standpoint, it could sometimes yield a more useful result.

Another measure to consider is the coefficient of forecast efficiency (5, p. 178). Most statistical textbooks describe the difference between the coefficient of correlation and the coefficient of determination where the latter is a squared version of the former, but they sometimes fail to call attention to the coefficient of forecast efficiency which has been designed to explain the predictive efficiency of a given correlation coefficient. The coefficient of forecast efficiency (E) is based on the coefficient of alienation which in itself is a measure designed to show the absence of relationship between two variables.[6]

This coefficient of forecast efficiency (E) is calculated by subtracting the coefficient of alienation from 1, as indicated by the following formula:

$$E = 1 - \sqrt{1 - r^2}$$

It is based upon the standard error of estimate, and it shows to what extent a prediction is improved if the variables in the correlation are used rather than the mean of the dependent variable for all estimates. Since it is based on squared deviations and considers the square root of the coefficient of alienation, the coefficient of forecast efficiency might be a more practical measure than either the coefficient of correlation or the coefficient of determination. The three measures are compared in table 3. The E coefficient reflects more nearly the relationship which is explained by actual deviations rather than squared deviations. For instance, in the example previously cited the coefficient of forecast efficiency is 0.50, which is remarkably close to the c value of 0.49 calculated on the basis of actual deviations from the least squares line. Thus, even though the E coefficient only approximates the actual efficiency of the independent variables in explaining unsquared deviations, it is computed in terms of squared data which makes it advantageous for use in conjunction with traditional regression and correlation analysis.

In the final analysis, it is only when research results are disseminated to others that anything worthwhile can be achieved. This is a matter of communication, and communication must take place with nonprofessional as well as professional groups. These people not only need to know what the results are, but also how they were obtained. Presenting research findings to the layman or the nonmathematical economist can be a real problem when the research has been based on more advanced analytical techniques. In economics and the social sciences the necessity of making allowances for changing conditions makes it even more imperative that the uninitiated user of research findings be able to understand the methods used. As Stigler aptly put it in reference to mathematical economics,

Table 3.--Comparative values for the coefficient of correlation, coefficient of determination, and the coefficient of forecast efficiency

| r | $r^2$ | E |
|---|---|---|
| 1.00 | 1.00 | 1.00 |
| .90 | .81 | .56 |
| .80 | .64 | .40 |
| .70 | .49 | .29 |
| .60 | .36 | .20 |
| .50 | .25 | .13 |
| .40 | .16 | .08 |
| .30 | .09 | .05 |
| .20 | .04 | .02 |
| .10 | .01 | .005 |

[5] This is not the first time these ideas have been considered. See, for example, Gauss and Fechner's work in the early 1800's (14, pp. 83-85), and some of Yule's later work on the association of attributes (1897) (14, pp. 125-131). Thorndike and Spearman also did substantial work on this in the early 1900's (14, p. 136).

[6] Technically, the coefficient of alienation $1 - r^2$ indicates the extent to which the relationship departs from a perfect correlation.

These methods make it necessary to disregard signs, but they do provide a workable solution which could have a considerable advantage over the traditional method. They also allow the possibility of using the median rather than the mean as the base point from which to measure deviations. Since the median is the middle point, it has the useful property of being that point around which the sum of the absolute deviations is minimized.[5] Although the median is not as stable as the mean from a mathematical standpoint, it could sometimes yield a more useful result.

Another measure to consider is the coefficient of forecast efficiency (5, p. 178). Most statistical textbooks describe the difference between the coefficient of correlation and the coefficient of determination where the latter is a squared version of the former, but they sometimes fail to call attention to the coefficient of forecast efficiency which has been designed to explain the predictive efficiency of a given correlation coefficient. The coefficient of forecast efficiency (E) is based on the coefficient of alienation which in itself is a measure designed to show the absence of relationship between two variables.[6]

This coefficient of forecast efficiency (E) is calculated by subtracting the coefficient of alienation from 1, as indicated by the following formula:

$$E = 1 - \sqrt{1 - r^2}$$

It is based upon the standard error of estimate, and it shows to what extent a prediction is improved if the variables in the correlation are used rather than the mean of the dependent variable for all estimates. Since it is based on squared deviations and considers the square root of the coefficient of alienation, the coefficient of forecast efficiency might be a more practical measure than either the coefficient of correlation or the coefficient of determination. The three measures are compared in table 3. The E coefficient reflects more nearly the relationship which is explained by actual deviations rather than squared deviations. For instance, in the example previously cited the coefficient of forecast efficiency is 0.50, which is remarkably close to the c value of 0.49 calculated on the basis of actual deviations from the least squares line. Thus, even though the E coefficient only approximates the actual efficiency of the independent variables in explaining unsquared deviations, it is computed in terms of squared data which makes it advantageous for use in conjunction with traditional regression and correlation analysis.

In the final analysis, it is only when research results are disseminated to others that anything worthwhile can be achieved. This is a matter of communication, and communication must take place with nonprofessional as well as professional groups. These people not only need to know what the results are, but also how they were obtained. Presenting research findings to the layman or the nonmathematical economist can be a real problem when the research has been based on more advanced analytical techniques. In economics and the social sciences the necessity of making allowances for changing conditions makes it even more imperative that the uninitiated user of research findings be able to understand the methods used. As Stigler aptly put it in reference to mathematical economics,

Table 3.--Comparative values for the coefficient of correlation, coefficient of determination, and the coefficient of forecast efficiency

| r | $r^2$ | E |
|---|---|---|
| 1.00 | 1.00 | 1.00 |
| .90 | .81 | .56 |
| .80 | .64 | .40 |
| .70 | .49 | .29 |
| .60 | .36 | .20 |
| .50 | .25 | .13 |
| .40 | .16 | .08 |
| .30 | .09 | .05 |
| .20 | .04 | .02 |
| .10 | .01 | .005 |

[5] This is not the first time these ideas have been considered. See, for example, Gauss and Fechner's work in the early 1800's (14, pp. 83-85), and some of Yule's later work on the association of attributes (1897) (14, pp. 125-131). Thorndike and Spearman also did substantial work on this in the early 1900's (14, p. 136).

[6] Technically, the coefficient of alienation $1 - r^2$ indicates the extent to which the relationship departs from a perfect correlation.

"from the viewpoint of the profession, the translation (of research results) is absolutely necessary, not merely desirable . . . If the mathematical economist's results are suggestive or useful, these people have a right to know them. If the results are tentative and conjectural, these people have a right to test them. It is the fundamental obligation of the scholar to submit his results and methods to the critical scrutiny of his competent colleagues in a comprehensible fashion" (12, p. 37).

Thus, as researchers we need to think in terms of the basic problems that need to be solved and adapt our methods accordingly. Any given method should be used, but only where it is appropriate and preferably where the results are easily understood by those concerned with the problem. With this kind of philosophy we can expect a wider acceptance of our research results.

## Literature Cited


(1) Ashar, V. G., and T. D. Wallace. "A sampling study of minimum absolute deviations estimators." Oper. Res., Vol. 2, No. 5, Oct. 1963.

(2) David, F. N., and J. Neyman. "Extension of the Markoff theorem on least squares." Statis. Res. Memoirs, Vol. 2, Univ. London, Dec. 1938.

(3) Ezekiel, Mordecai, and Karl A. Fox. Methods of Correlation and Regression Analysis. John Wiley and Sons, New York, 3rd ed., 1959.

(4) Fisher, Ronald A. Statistical Methods for Research Workers. Hafner Publishing Co., New York, 13th ed., rev. 1958.

(5) Garrett, H. E. Statistics in Psychology and Education. David McKay Co., Inc., New York, 6th ed., 1966.

(6) Hart, William L. College Algebra. D.C. Heath and Co., Boston, 3rd ed., 1947.

(7) Havlicek, J. "Use of linear programming for estimating regression relationships." Proc. Assoc. Southern Agr. Workers, Atlanta, Ga., Feb. 1964.

(8) Hume, David. "An inquiry concerning human understanding." The English Philosophers from Bacon to Mill. E. A. Burtt, ed., Modern Library, New York, 1939.

(9) Moroney, M. J. Facts from Figures. Penguin Books Ltd., Great Britain, 3rd ed., 1956.

(10) Snedecor, George W. Statistical Methods. The Iowa State College Press, Ames, 4th ed., 1946.

(11) Steel, R. G. D., and J. H. Torrie. Principles and Procedures of Statistics. McGraw-Hill Book Co., New York, 1960.

(12) Stigler, George J. "The mathematical method in economics." Five Lectures on Economic Problems. London Sch. Econ. and Polit. Sci., Macmillan Co., New York, 1950.

(13) Wagner, H. M. "Linear programming techniques for regression analysis." Jour. Amer. Statis. Assoc., Vol. 54, 1959.

(14) Walker, Helen M. Studies in the History of Statistical Method. The Williams and Wilkins Co., Baltimore, Md., 1929.

(15) Waugh, Albert E. Elements of Statistical Method. McGraw-Hill Book Co., New York, 3rd ed., 1952.