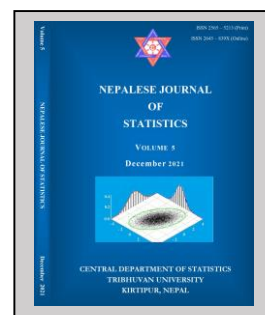# A Note on Confusion between Linear and Affine Functions and the Generalized Forms of Gradient

Supriya Malla[1] and Ganesh Malla[2*]

ABSTRACT

Background: Arguably the most frequently used term in science, particularly in mathematics and statistics, is *linear*. However, confusion arises from the various meanings of linearity instructed in different levels of mathematical courses. The definition of linearity taught in high school is less correct than the one learned in a linear algebra class. The correlation coefficient of two quantitative variables is a numerical measure of the affinity, not only linearity, of two variables. However, every statistics book loosely says it is a measure of linear relationship. This clearly show that there is some confusion between use of the terms the linear function and affine function.

Objective: This article aims at clarifying the confusion between use of the terms linear function and affine function. It also provides more generalized forms of the gradient in different branches of mathematics and show their equivalency.

Materials and Methods: We have used the pure analytical deductive methods to proof the statements.

Results: We have clearly presented that gradient is the measure of affinity, not just linearity. It becomes a special case of the derivative in calculus, of the least-squares estimate of the regression coefficient in statistics and matrix theory. The gradient can be seen in terms of the inverse of the informative matrix in the most general setting of the linear model estimation.

Conclusion: The article has been clearly written to show the distinction between the linear and affine functions in a concise and unambiguous manner. We hope that readers will clearly see various generalizations of the gradient and article itself would be a simple exposition, enlightening, and fun to read.

Keywords: Affine function, derivative, inverse informative matrix, least-squares estimate, regression coefficient.

Address correspondence to the author: Cornell University, Ithaca, NY 14850, USA.
Email: sm2427@cornell.edu[1];
University of Cincinnati-Clermont College, Batavia, OH, 45103, USA.
Email: ganesh.malla@uc.edu[2*] (Corresponding author email)

INTRODUCTION

Studying the term 'linear' to a scientist is similar to 'finding God in all things' for a religious person. Linear equations, linear algebra, linear regression, linear interpolation, linear approximation, linear differential equations, linear discriminant analysis, linear programming, linear inequality, linear independence, linear momentum, linear optimization, linear operator, and linear kinetics are just a few topics that are associated with the term linear. Many great minds have dedicated their lives to study just one or two of these fields. The popularity of the word itself is so great that even innovators retain the term in their products' names: linear actuator (a tool that creates motion in a straight line and is used in various machines like computers, hydraulic cylinder etc.), linear garage door opener, linear jet, linear motor, linear software, linear technology, linear wireless, etc. Some words that are synonymous, although in various senses, with the word linear are collinear, line-like, lineal (not multidimensional), and simple (not compounded). Nonlinear is a very general term used for linear antonym.

The website http://www.synonym.com/definition/linear, and Stewart et al. (2012) define a linear function as a function designating or involving an expression whose terms are of the first degree. That is, if the dependent (response) variable f is a function of a single variable x, then the proceeding definition of a linear function can be replaced by:

$f(x) = mx + c$, where m and c are constant, $m \neq 0$ ………………………. (1)

The definition of a linear function in linear algebra

In Lay et al. (2015) and the universally accepted definition of a linear function is not in coherence with the definition in (1). Before we prove this, let us first see the definition of a linear function in linear algebra. Let V and W be vector spaces. A function $f: V \rightarrow W$ is said to be a *linear function* if for any two vectors x and y in V and any scalars α and b in the real number system satisfy the condition:

$f(ax + by) = a\, f(x) + b\, f(y)$ ………………………………………………… (2)

Now, we want to show that definition (1) contradicts definition (2). To verify this, let the function f have the form $f(x) = mx + c$. Then, for any two arguments x and y in V and a and b two real numbers,

LS of (2)   = f(ax+ by)
            = m(ax+ by) + c
            = a (mx +c) + b (my + c) + c(1-a-b)
            = a f(x) + b f(y) + c (1-a-b) $\neq$ RS of (2).

Now, what this result tells us is that the definition of a linear function in linear algebra envisages the fact that when the argument is zero, the function must be zero. However, definition (1) does not bother with this. The definition of a linear function in linear algebra is analogous with the Japanese system of measuring age which begins not when a newborn leaves his mother's womb, rather when he is first conceived.

In Lay et al. (2015), a function f: V → W in the setting for defining the equation (2) with d in W is said to be an a*ffine function* if:

f(ax + by) = a f(x) + b f(y) + d …………………………………….. (3)

The most general definition of a linear function

Let A be a matrix of order m x n and X be a vector from R^n. Then, a function from f: $R^m \to$ R^n is said to be a linear function if

f(X) = AX …………………………………………………….. (4)

(Rencher, 2008). It is not difficult to show that this definition of a linear function agrees with the definition of a linear function in linear algebra.

Some generalized forms of slope

Although linear and affine functions differ (an affine function is a linear function with a translation), the main measurement of association between the response variable (y) and explanatory variable (x) is slope. The slope (defined as the ratio of rise to run) in even an elementary level text becomes an important measurement. The same term, slope, is called the derivative in calculus; the coefficient of regression of y w. r. t. x in the least squares regression line of y with x; the mortality rate in epidemiology; the force of market in economics, and so on and so forth.

The calculus generalization of the slope

The concepts of limit and derivative of a function in Mendelson (1995) are arguably the two most important concepts of calculus as they form the basis of calculus and hence, much of mathematics. It is interesting to realize the fact that the concept of derivative has solely been defined to deal with the slope of tangent lines at different points. The derivative rule $f'(x) = n\,x^{n-1}$ when $f(x) = x^n$ states the fact that $f'(x) = m$, the slope of the line when f(x) = mx + c.

The statistical generalization of the slope

When the relationship between two quantitative variables is imperfect or probabilistic in nature, statisticians still like to know what is the change in y (on average) when the explanatory variable is increased by a unit. In 1794, Carl Friedrich Gauss described a method called the least-squares method which has become an important tool in data modeling and aids in finding the slope between two variables. The principle of least-squares (PLS) states that fitting a model to the data minimizes the sum of squared errors—an error being the difference between an observed value and the fitted value provided by a model. What could be a better way to do something by minimizing errors, in fact, by minimizing the sum of the squares of the errors? For this reason, perhaps, the model one gets by using this method is called the model of best fit! The estimate of the parameter b of the theoretical line y = b x + c using the PLS to the data set {(x₁, y₁), (x₂, y₂), … , (xₙ, yₙ)} is given by:

$$\widehat{b} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2} \quad \text{……………………………………………………(5)}$$

Now, we want to show that the least square estimate $\hat{b}$ is exactly the slope when there is a perfect line relationship. In that case, $y_i = b\, x_i + c$, for i = 1, 2, 3, …, n.

Next, if all the points lie on a straight line $y_i = m\, x_i + c$, then we can pick two points (x_1, y_1) and (x_2, y_2) on the straight line and thus if $y_i = m\, x_i + c$, then

$$\bar{y} = m\bar{x} + c.$$

Then, $\hat{b} = \dfrac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$

$$= \dfrac{\sum_i^n (x_i - \bar{x})(mx_i + c - m\bar{x} - c)}{\sum_i^n (x_i - \bar{x})^2}$$

$$= m\, \dfrac{\sum_i^n (x_i - \bar{x})^2}{\sum_i^n (x_i - \bar{x})^2}$$

$= m$, which is the slope of the line.

The generalization of slope in matrix form of data

The Sample Linear Regression model in matrix form, in Rencher (2008) [4], is

$$y_{n \times 1} = \beta x + \varepsilon_{n \times 1},$$

where $y = [y_1\ y_2\ …\ y_n]'$ and $x = [x_1\ x_2\ …\ x_n]'$. The vector y is called the response vector; x the design vector; ε the error vector, and β the parameter vector of unknowns.

Under the assumption that ε follows $N(0, \sigma^2 I)$ where I is an $n \times n$ identity matrix, by using the PLS, that is minimizing the sum of the squares of the error:

$$\varepsilon'\epsilon = (y - \beta x)'(y - \beta x),$$

one gets the estimate of β,

$$\hat{\beta} = (x'x)^{-1}x'y.$$

Now, in reality, this is just to convince you that it is nothing but the generalized form of the slope in matrix form. However, we want to prove it. If there is $y_i = m\, x_i$, i = 1, 2, …, n, a proper relationship, then

$$\hat{\beta} = [(x_1, x_2, … , x_n)(x_1, x_2, … , x_n)']^{-1}[(x_1, x_2, … , x_n)(y_1, y_2, … , y_n)']$$

$$= (x_1^2 + x_2^2 + … + x_n^2)^{-1}(x_1 y_1 + x_2 y_2 + \cdots + x_n y_n)$$

$$= (x_1^2 + x_2^2 + … + x_n^2)^{-1} m(x_1^2 + x_2^2 + … + x_n^2)$$

$= m$, which is the slope.

CONCLUSION

The article clearly articulates the clear distinctions between the linear and affine functions mathematically and hope to eliminate the incorrect use of the term linear in most of the elementary mathematics texts. We recommend that all texts clearly write f(x) = mx + b, b ≠ 0, is 'not' a linear function, rather an affine function and only f(x) = mx is a linear function. We also discover and present how the math concept slope does take many generalized forms with different names in higher mathematics like calculus, statistics, and matrix theory, which is very interesting to notice.

CONFLICT OF INTEREST

The authors declared that there is no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

Stewart, J., Redlin, L., & Watson, S. (2012). *Algebra and Trigonometry* (4th ed.). USA: The Cengage Learning.

Mendelson, E. (1995). *Schaum's Outline of Beginning Calculus* (3th ed.). USA: The McGraw-Hill Companies, Inc.

Lay, D., Lay, S., & McDonald, J. (2015). *Linear Algebra and Its Applications* (5th ed.). USA: Pearson Pub.

Rencher, A. (2008). *Linear Models in Statistics* (2nd ed.). USA: John Wiley & Sons, Inc.