Journal of Probability and Statistical Science
21(1), 133-134, Feb. 2023

Noncalculus Derivations of Least Squares Estimates

Lawrence M. Lesser
The University of Texas at El Paso

ABSTRACT
This letter responds to and reflects on goals and claims of the previous issue's papers on deriving least-squares estimates without calculus and brings in the technique of translating data points.

Keywords: *teaching, regression, least squares*

How interesting that consecutive papers [1,9] from the August 2022 issue of *JPSS* focus on deriving least-squares estimates without calculus, an admirable goal because statistics is being taken by an increasing number of students with little or no background in calculus. More generally, [2] delineates different approaches specific textbooks take in how students can derive or understand the formulas for least squares estimates: 'black box' (using technology or formulas without justification), differential calculus, vector projection (for linear algebra or advanced statistics texts), sequential completing the square in two variables, and then (by leveraging the line going through the centroid) completing the square in only one variable.

As an aside, I agree that the broader goal of demystifying least-squares regression is quite worthy, and I would add that it should not be limited to the formulas themselves. Classes should also discuss the motivation for the least-squares criteria in the first place. I offer a natural classroom sequence [3] of questions for exploring this: Why not use *perpendicular* deviations? Why not minimize the *sum* of the vertical deviations? Why not minimize the sum of the *absolute* deviations? Why minimize the sum of the *squared* deviations?

Sarkar and Rashid [9] offer two noncalculus derivations of the least-squares formulas. The paper's Method 1 uses a minimization technique that does not require calculus because it uses the high school algebra technique of completing the square, needing to do it only in a single variable by leveraging the fact that the general least squares line passes through the centroid ($\bar{x}$, $\bar{y}$). A derivation with this idea has been presented in, for example, a college algebra textbook [8].

What is also worth noting is the strategy [5,6,7] of translating all ordered pairs of data by the centroid (or even translating by only the x-coordinates), performing the least squares procedure on the translated points, and then translating the result back to the original scale. The beauty of this technique is that it does not depend on the property of the line passing through the centroid which means it needs to complete the square in *two* variables, but because the translation eliminates the mixed term (i.e., a multiple of the product of the variables) from the expression, completing the square can be readily done in each variable separately within the same expression.

The idea of translating data points should not seem strange to someone knowing high school algebra, given that its flavor is not unlike the technique of variable substitution that is sometimes used to simplify and/or solve algebra equations. For example, a student might use the substitution $y = x^3$ to change the sixth-degree equation $x^6 - 4x^3 - 21 = 0$ into a quadratic equation

Received November 2022, in final form January 2023.
□ Lawrence M. Lesser (corresponding author) is affiliated with the Department of Mathematical Sciences, The University of Texas at El Paso

in y. Or a student might use the formula $x = y - \frac{b}{2a}$ to translate the quadratic equation $ax^2 + bx + c = 0$ into a quadratic equation in y that can be solved without needing factoring or the quadratic formula because there is no longer a linear term. And, of course, a statistics student is used to translations by standardizing data into z-scores.

The paper's Method 2 takes a weighted mean of slopes from all possible pairs of the n points from the dataset, and it is claimed that "high school algebra suffices." While it is true no calculus technique is needed, the mathematical maturity needed to follow the argument, the choice of weights, and its notation may not be readily grasped by those whose math background does not go beyond high school algebra.

Also, Method 2 relies on the best-fitted line passing through the centroid of the data. However, there are many real-world situations [4] that call for a zero-intercept regression model and there are also mathematical advantages (even in more advanced courses [1]) to beginning with the zero-intercept case. Therefore, it would be important to note that the best-fitted line actually does NOT pass through the centroid in the zero-intercept case.

References

[1]. Inlow, M. (2022). On Deriving the Least Squares Estimates in Introductory Regression Courses, *Journal of Probability and Statistical Science*, **20**(1), 184-190.

[2]. Key, E. S. (2005). A Painless Approach to Least Squares, *The College Mathematics Journal*, **36**(1), 65- 67.

[3]. Lesser, L. (1999a). The 'Ys' and 'Why Nots' of Line of Best Fit, *Teaching Statistics*, **21**(2), 54-55.

[4]. Lesser, L. (1999b). Making the Black Box Transparent, *Mathematics Teacher*, **92**(9), 780-784.

[5]. Lesser, L. (2004). Translated Least Squares, *Mathematics Teacher*, **98**(2), 68-69.

[6]. Lesser, L. (2008). More on Least Squares, *Mathematics Teacher*, **102**(1), 6-7.

[7]. Lesser, L. (2017). Completing the (Least) Squares, *Mathematics Teacher*, **111**(2), 88.

[8]. Mayes, R. and Lesser, L. (1998). *ACT in Algebra: Applications, Concepts and Technology in Learning Algebra* (Preliminary ed.). Burr Ridge, IL: McGraw-Hill.

[9]. Sarkar, J. and Rashid, M. (2022). Estimating the Parameters of a Simple Linear Regression Model Without Using Differential Calculus, *Journal of Probability and Statistical Science*, **20**(1), 191-203.