Open Journal of Statistics, 2025, 15(3), 251-271
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Fixed Points in Linear Regression

David L. Farnsworth

School of Mathematics and Statistics, Rochester Institute of Technology, Rochester, New York, USA
Email: dlfsma@rit.edu

How to cite this paper: Farnsworth, D.L. (2025) Fixed Points in Linear Regression. *Open Journal of Statistics*, **15**, 251-271.
https://doi.org/10.4236/ojs.2025.153013

Received: April 19, 2025
Accepted: June 6, 2025
Published: June 9, 2025

Abstract

There is a set of points in the plane whose elements correspond to the observations that are used to generate a simple least-squares regression line. Each value of the independent variable in the observations matches up with one of these points, which are called pivot or fixed points. The coordinates of the fixed points are derived, and the properties of the points are explored. All points in the plane that yield each of the fixed points are found. The role that fixed points play in regression diagnostics is investigated. A new mechanical device that uses linkages to model the role of fixed points is described. A numerical example is presented.

Keywords

Fixed Point, Pivot Point, Influence, Leverage, Regression Diagnostic, Mechanical Device

1. Introduction

It is shown in [1] that each point in a set of observations that are used in a standard least-squares linear fit has an accompanying *pivot point* or *fixed point*. If a selected data point is replaced by any point on the vertical line containing the selected point, the new regression line, which is based upon the other original points and the new point on the vertical line, contains the pivot point, i.e., all these regression lines are concurrent, including the one for the original points. Varying the observation in this manner in the direction of prediction, which is parallel to the vertical axis for simple linear regression, is a common technique for diagnosing each observation's impact on the regression line [1] ([2], p. 321) ([3], pp. 384-420) [4].

Figure 1 illustrates this phenomenon for a set of artificial observations. The observation R, which is on the dashed vertical line $x = 7$, has been selected and consecutively replaced by one other point on the vertical line. The other points plus a replacement point on the vertical line produce the displayed regression lines. All of the regression lines contain P, which is the pivot or fixed point corresponding to observation R. There is a regression line for each point on $x = 7$. This graph is related to the numerical example in Section 6.

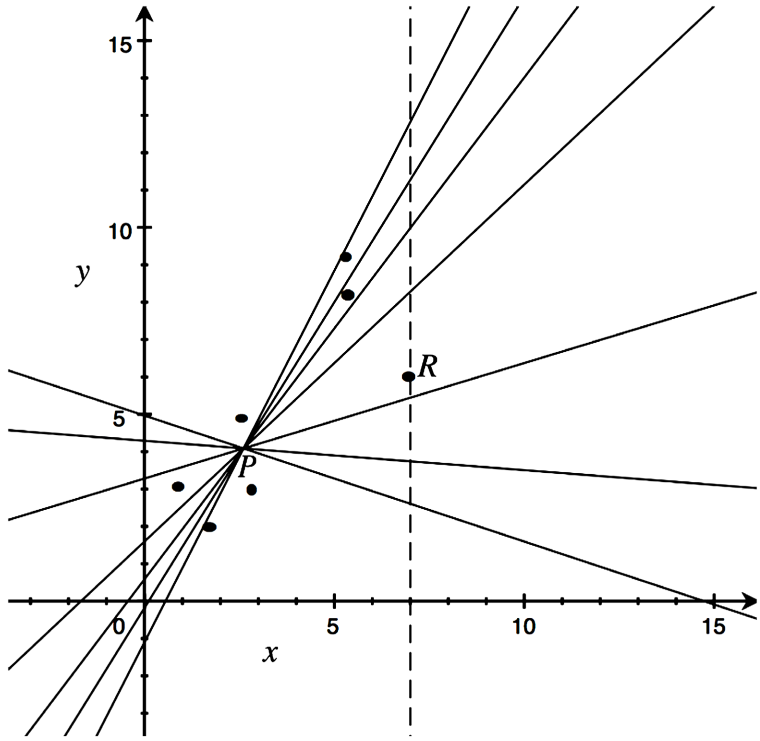

Figure 1. The pivot or fixed point P corresponding to the observation R.

Although figures similar to **Figure 1** had appeared, the existence of the pivot or fixed point was not pointed out until recently [1]. For example, a figure in ([4]: p. 390) contains an unnoted point on five of these regression lines for the main data set of that article. In ([5]: p. 381) and ([6]: p. 98), the authors use identical graphs to discuss the large effect that one bivariate observation can have, but they do not note the pivot point that is in the graphs.

In Section 2, the existence of pivot points is proven and their coordinates are found. Then, those coordinates are expressed differently in Section 3 in order to show the properties of pivot points. In Section 4, a more general approach is taken to find all points in the plane that can replace the selected point to create a regression line containing the originally selected point's pivot point. The geometry of these special points is explored in Section 5. A numerical example is supplied in Section 6, followed by concluding comments in Section 7.

2. Existence of Pivot or Fixed Points

Consider the least-squares regression line

$$y = \beta_0 + \beta_1 x + \varepsilon, \tag{1}$$

where the intercept β_0 and slope β_1 are to be estimated and the error term ε has a standard normal distribution. The observations are

$$S = \{(x_i, y_i) \mid i = 1, 2, \cdots, n\}, \tag{2}$$

where not all x-values are equal. Recall that in order to obtain estimates of the y-intercept and slope, which are the outcomes of the minimization of the sum of squared vertical distances of the data from the regression line, apply $\frac{\partial}{\partial \beta_0}$ and $\frac{\partial}{\partial \beta_1}$ to

$$\sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2. \tag{3}$$

Setting the results equal to zero and introducing the estimates b_0 and b_1 of β_0 and β_1, respectively, give the Normal Equations

$$nb_0 + \left(\sum_{i=1}^{n} x_i\right) b_1 = \sum_{i=1}^{n} y_i \tag{4}$$

and

$$\left(\sum_{i=1}^{n} x_i\right) b_0 + \left(\sum_{i=1}^{n} x_i^2\right) b_1 = \sum_{i=1}^{n} x_i y_i \tag{5}$$

([2], p. 305) ([3], p. 17) Differentiation can be avoided by noticing that the sum (3) can be written, using the estimates, as

$$\begin{aligned} \sum_{i=1}^{n} \left(y_i - (b_0 + b_1 x_i)\right)^2 = {} & \sum_{i=1}^{n} y_i^2 + nb_0^2 + \sum_{i=1}^{n} x_i^2 b_1^2 - 2\sum_{i=1}^{n} y_i b_0 \\ & + 2\sum_{i=1}^{n} x_i b_0 b_1 - 2\sum_{i=1}^{n} x_i y_i b_1, \end{aligned}$$

which is separately quadratic in b_0 and b_1 with positive leading coefficients n and $\sum_{i=1}^{n} x_i^2$, respectively. As a quadratic expression in b_0, the minimum is attained at

$$b_0 = -\frac{2b_1 \sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} y_i}{2n},$$

which is Equation (4). As a quadratic expression in b_1, the minimum is attained at

$$b_1 = -\frac{2b_0 \sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i}{2\sum_{i=1}^{n} x_i^2},$$

which is Equation (5).

Dividing the Normal Equation (4) by sample size n shows that the point of means

$$\left(\sum_{i=1}^{n} x_i / n, \sum_{i=1}^{n} y_i / n\right) \tag{6}$$

is on the regression line in Equation (1).

Solving the Normal Equations (4) and (5) gives

$$b_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{7}$$

and

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \tag{8}$$

Thus, the regression line in Equation (1) can be written as

$$y = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} + \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} x, \tag{9}$$

and as

$$y = \frac{\sum_{i=1}^{n} y_i}{n} + \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\left(x - \frac{\sum_{i=1}^{n} x_i}{n}\right). \tag{10}$$

A fixed point or pivot point in simple linear regression is obtained in the following way. Given a set of bivariate observations, select one of the points, whose y-value will be made to vary, while leaving unchanged all the observations' x-values and all their y-values except the selected one. Repeatedly, find the regression line for the original observations but with a replacement y-value for the selected one. All the regression lines created in this manner contain one point, which is called the pivot point in [1]. The regression lines pivot about the point that is like a fulcrum for a lever, as in **Figure 1**. The lever in various positions is analogous to the regression lines. The term fixed point is nearly exclusively used below for this point. Theorem 1 says that each observation in a bivariate data set is accompanied by a fixed point.

Equations (7) and (8) show the well-known linear dependence of b_0 and b_1, and, hence, the regression line in Equations (9) and (10), on each y-value among the observations ([3], pp. 18-19) ([4], p. 390). Designate by y_n the selected y-value that is to be replaced in order to investigate the fixed point. The observations can be ordered so that the data point of interest is the nth one having coordinates (x_n, y_n). Equations (7) and (8) are written to explicitly display the linear dependence of the coefficients on y_n as

$$b_0 = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n-1} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n-1} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} + \frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right) x_n}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} y_n \tag{11}$$

and

$$b_1 = \frac{n\sum_{i=1}^{n-1} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n-1} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} + \frac{n x_n - \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} y_n. \tag{12}$$

Theorem 1 (Existence and Coordinates of the Fixed Point in Simple Linear

Regression). *Consider the set of bivariate observations* S *in Equation* (2), *where not all* x*-values are equal. Select* $R(x_n, y_n)$, *where* x_n *is not the sample mean, i.e.,* $x_n \neq \overline{x}_{(n)}$. *For these data, each value* t *that is chosen to replace* y_n *produces a member of a family of least-squares regression lines. All members of this family contain one point* P, *the fixed point, which is on the regression line for* S. *In terms of the first* $n-1$ *observations' coordinates and* x_n, P *is*

$$P(p,q) = P\left(\overline{x}_{(n)} - \frac{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}{n\left(x_n - \overline{x}_{(n)}\right)}, \frac{n-1}{n}\overline{y}_{(n-1)} - \frac{\sum_{i=1}^{n-1}\left(x_i - \overline{x}_{(n)}\right)y_i}{n\left(x_n - \overline{x}_{(n)}\right)}\right). \quad (13)$$

If $x_n = \overline{x}_{(n)}$, *then it is said that the observation has a fixed point and that the point is at infinity. Subscripts in parentheses indicate sample size. In particular,* $\overline{x}_{(n)}$ *is the sample mean of all* n x*-values in* S, *and* $\overline{y}_{(n-1)}$ *is the sample mean of the* y*-values for the first* $n-1$ *observations of* S.

Proof 1. The slope in Equation (9) and (10) can be written in a standard way as

$$\begin{aligned} b_1 &= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{n\sum_{i=1}^{n}\left(x_i - \overline{x}_{(n)}\right)y_i}{n\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2} \\ &= \sum_{i=1}^{n}\left(\frac{x_i - \overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}\right)y_i \end{aligned} \quad (14)$$

([2]: p. 306) ([3]: pp. 17-18) Recalling that the point of means in Formula (6) is on the line, the regression line for S can be written so that the coefficient of y_n is isolated as

$$\begin{aligned} y &= \overline{y}_{(n)} + b_1\left(x - \overline{x}_{(n)}\right) \\ &= \frac{\sum_{i=1}^{n} y_i}{n} + \sum_{i=1}^{n}\left(\frac{x_i - \overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}\right)y_i\left(x - \overline{x}_{(n)}\right) \\ &= \left(\frac{\sum_{i=1}^{n-1} y_i}{n} + \sum_{i=1}^{n-1}\left(\frac{x_i - \overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}\right)y_i\left(x - \overline{x}_{(n)}\right)\right) \\ &\quad + \left(\frac{1}{n} + \frac{\left(x_n - \overline{x}_{(n)}\right)\left(x - \overline{x}_{(n)}\right)}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}\right)y_n. \end{aligned} \quad (15)$$

Consider an arbitrary point $T(x_n, t)$ on the vertical line

$$x = x_n \quad (16)$$

as a temporary replacement for the observation $R(x_n, y_n)$. See **Figure 2**. Then, Equation (15) becomes

$$y=\left(\frac{\sum_{i=1}^{n-1}y_i}{n}+\sum_{i=1}^{n-1}\left(\frac{x_i-\overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}\right)y_i\left(x-\overline{x}_{(n)}\right)\right)$$
$$+\left(\frac{1}{n}+\frac{\left(x_n-\overline{x}_{(n)}\right)\left(x-\overline{x}_{(n)}\right)}{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}\right)t. \tag{17}$$

If x in Equation (17) is the x-coordinate p of a fixed point P, then the y-coordinate q of P needs to be independent of t, and thus the coefficient of t in Equation (17) must be zero, i.e.,

$$\frac{1}{n}+\frac{\left(x_n-\overline{x}_{(n)}\right)\left(x-\overline{x}_{(n)}\right)}{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}=0.$$

Thus,

$$x=\overline{x}_{(n)}-\frac{\sum_{i=1}^{n}\left(x_i-\overline{x}_{(n)}\right)^2}{n\left(x_n-\overline{x}_{(n)}\right)}, \tag{18}$$

which is p in Formula (13). By substituting the x-Coordinate (18) into Equation (17),

$$\begin{aligned}y&=\frac{\sum_{i=1}^{n-1}y_i}{n}+\sum_{i=1}^{n-1}\left(\frac{x_i-\overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}\right)y_i\left(x-\overline{x}_{(n)}\right)\\&=\frac{n-1}{n}\overline{y}_{(n-1)}+\sum_{i=1}^{n-1}\left(\frac{x_i-\overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}\right)y_i\left(-\frac{\sum_{j=1}^{n}\left(x_j-\overline{x}_{(n)}\right)^2}{n\left(x_n-\overline{x}_{(n)}\right)}\right)\\&=\frac{n-1}{n}\overline{y}_{(n-1)}-\frac{1}{n}\sum_{i=1}^{n-1}\left(\frac{x_i-\overline{x}_{(n)}}{x_n-\overline{x}_{(n)}}\right)y_{i,}\end{aligned}$$

which is q in Formula (13). ■

Proof 2. All members of the family of lines

$$y=\left(a+cr\right)+\left(d+er\right)x \tag{19}$$

with real parameter r and real numbers a,c,d and e for $e\neq 0$ intersect at the point

$$\left(-\frac{c}{e},\frac{ae-cd}{e}\right). \tag{20}$$

The validity of Point (20) is checked by substitution for x and y from Point (20) into Equation (19). If $e=0$, then the lines in Equation (19) are parallel with slope d and, thus, do not intersect, unless they coincide, which occurs if, additionally, $c=0$. The existence of the fixed point can be seen from the identification of the coefficients in Formulas (11) and (12) with the corresponding coefficients in Equation (19) for $r=y_n$. The coordinates of P can be obtained by substituting from Formulas (11) and (12) for a, c, d and e into the coordinates of the point in (20). Comparing Formulas (11) and (12) with Equation (19), $e \neq 0$ is equivalent to $x_n \neq \overline{x}_{(n)}$. ■

Figure 2 displays the chosen observation $R(x_n, y_n)$, the vertical line $x = x_n$, one example of a replacement point $T(x_n, t)$, the regression line for the set $S \setminus \{(x_n, y_n)\} \cup T$, and the fixed point P. The observations are not displayed in the graph.

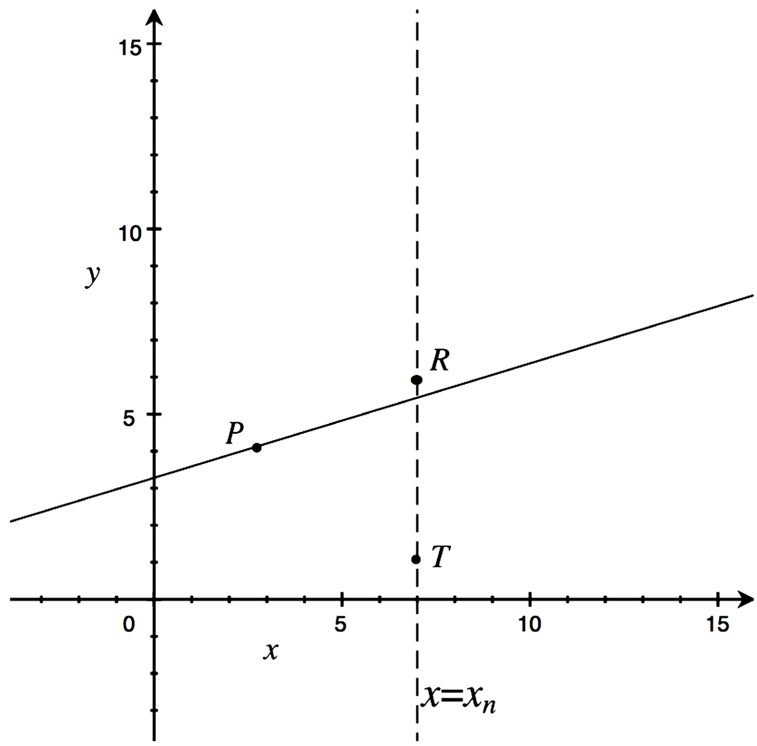

Figure 2. An example of a point $T(x_n, t)$ that replaces then selected observation R in Theorem 1.

3. Properties of the Fixed Point

The first goal of this section is to express the coordinates of the fixed point P so that they include y_n in a way similar to the inclusion of the x-coordinate x_n. The second goal is to write the coordinates of the fixed point so that the occurrences of x_n are explicit throughout and the coordinates are given in terms of the simplest sums. The advantage of these re-expressions is that they make some properties of fixed points much more apparent. The third goal is to show that the fixed point is the intersection of the original regression line for the set S and the regression line for the set of the first $n-1$ observations of S, *i.e.*, for $S \setminus \{(x_n, y_n)\}$.

3.1. Expression of the Coordinates of the Fixed Point to Include y_n

Although the fixed point's coordinates can be exhibited excluding the y-coordinate y_n of the selected point, as in Formula (13) of Theorem 1, the coordinates may be written to incorporate y_n in a way similar to the involvement of the x-coordinate x_n. The occurrences of y_n can be made to drop out algebraically, because the fixed point P does not depend upon y_n. It can be useful not to be limited to expressions free of y_n.

In the slope b_1 in Formula (14), designate

$$k_i = \frac{x_i - \overline{x}_{(n)}}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2},$$

which are the weighing factors of the observations' y-coordinates for computing the slope. Although the k_i do not appear to have a name, they occur often. Three of their properties are

$$\sum_{i=1}^{n} k_i = 0, \quad \sum_{i=1}^{n} x_i k_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} k_i^2 = \frac{1}{\sum_{j=1}^{n}\left(x_j - \overline{x}_{(n)}\right)^2}$$

([3], p. 42). The coordinates of the fixed point $P(p,q)$ in Formula (13) can be rewritten as follows. The x-coordinate is

$$p = \overline{x}_{(n)} - \frac{1}{nk_n}.$$

Because both the point of means in Formula (6) and the fixed point P are on the original regression line of slope b_1, the y-coordinate q satisfies

$$\frac{\overline{y}_{(n)} - q}{\overline{x}_{(n)} - p} = \frac{\overline{y}_{(n)} - q}{\overline{x}_{(n)} - \left(\overline{x}_{(n)} - \frac{1}{nk_n}\right)} = b_1.$$

Thus,

$$q = \overline{y}_{(n)} - \frac{b_1}{nk_n},$$

and the fixed point can be expressed as

$$P(p,q) = P\left(\overline{x}_{(n)} - \frac{1}{nk_n}, \overline{y}_{(n)} - \frac{b_1}{nk_n}\right). \tag{21}$$

This presents a balanced notation for the two coordinates and sets out the horizontal and vertical distances between P and the point of means. It shows that the fixed point can be far from the observations when k_n is small, i.e., when x_n is near $\overline{x}_{(n)}$, and for regression lines with slopes that are large in absolute value.

3.2. Expression of the Fixed Point with the Simplest Sums

The identities

$$n\left(x_n - \overline{x}_{(n)}\right) = (n-1)\left(x_n - \overline{x}_{(n-1)}\right), \tag{22}$$

$$\sum_{i=1}^{n}\left(x_i - \overline{x}_{(n)}\right)^2 = \sum_{i=1}^{n-1}\left(x_i - \overline{x}_{(n-1)}\right)^2 + \frac{n-1}{n}\left(x_n - \overline{x}_{(n-1)}\right)^2, \tag{23}$$

and

$$\sum_{i=1}^{n}\left(x_i-\overline{x}_{(n)}\right)y_i=\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)y_i+\frac{n-1}{n}\left(x_n-\overline{x}_{(n-1)}\right)\left(y_n-\overline{y}_{(n-1)}\right) \tag{24}$$

are algebraic. Using Identities (22) - (24), the fixed point in Formula (13) can be written as

$$P\left(p,q\right)=P\left(\overline{x}_{(n-1)}-\frac{\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)^2}{\left(n-1\right)\left(x_n-\overline{x}_{(n-1)}\right)},\overline{y}_{(n-1)}-\frac{\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)\left(y_i-\overline{y}_{(n-1)}\right)}{\left(n-1\right)\left(x_n-\overline{x}_{(n-1)}\right)}\right) \tag{25}$$

and

$$P\left(p,q\right)=P\left(\frac{\sum_{i=1}^{n-1}x_i}{n-1}-\frac{\left(n-1\right)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2}{\left(n-1\right)\left(\left(n-1\right)x_n-\sum_{i=1}^{n-1}x_i\right)},\frac{\sum_{i=1}^{n-1}y_i}{n-1}-\frac{\left(n-1\right)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i}{\left(n-1\right)\left(\left(n-1\right)x_n-\sum_{i=1}^{n-1}x_i\right)}\right). \tag{26}$$

These give the coordinates of P as sums involving x_n and the first $n-1$ observations. The form of the dependence upon x_n is made clearer. They illustrate that, as x_n approaches $\overline{x}_{(n-1)}$, the fixed point's coordinates increase without bound. Like in Formula (21), Equation (26) conveys the two coordinates of P in a way to reveal a congruity between them. In Formula (26), the coordinates are expressed by employing only the most rudimentary sums.

From Formula (25),

$$p-\overline{x}_{(n-1)}=-\frac{\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)^2}{\left(n-1\right)\left(x_n-\overline{x}_{(n-1)}\right)}, \tag{27}$$

and thus

$$x_n-\overline{x}_{(n-1)}=-\frac{\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)^2}{\left(n-1\right)\left(p-\overline{x}_{(n-1)}\right)}, \tag{28}$$

showing that the x-coordinate x_n of the selected point can be easily computed from the x-coordinate p of the fixed point. Additionally, Equations (27) and (28) show that x_n and p are the same function of each other through a function that otherwise depends upon only the first $n-1$ observations. Writing Equation (27) as

$$\left(x_n-\overline{x}_{(n-1)}\right)\left(p-\overline{x}_{(n-1)}\right)=-\frac{\sum_{i=1}^{n-1}\left(x_i-\overline{x}_{(n-1)}\right)^2}{n-1} \tag{29}$$

makes apparent the hyperbolic nature of the relationship between the nth observation's first coordinate x_n and its accompanying fixed point's first coordinate p. By considering x_n and p as coordinates, Equation (29) is a rectangular hyperbola whose axis has slope minus one and contains the center $\left(\overline{x}_{(n-1)},\overline{x}_{(n-1)}\right)$. The asymptotes are $x_n=\overline{x}_{(n-1)}$ and $p=\overline{x}_{(n-1)}$.

The y-coordinate q of the fixed point P produces x_n as well. One way to accomplish that is to find p from q using a regression line that contains $P(p,q)$. Another way is to employ the y-coordinate in Formula (25) to obtain

$$x_n = \bar{x}_{(n-1)} - \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)\left(y_i - \bar{y}_{(n-1)}\right)}{(n-1)\left(q - \bar{y}_{(n-1)}\right)}. \tag{30}$$

This shows that $x_n - \bar{x}_{(n-1)}$ and $q - \bar{y}_{(n-1)}$ are the same function of each other.

3.3. The Fixed Point as the Intersection of Two Specific Lines

Theorem 2 gives the location of a fixed point as the intersection of two particular regression lines.

Theorem 2 (The Identity of the Fixed Point as a Particular Point of Intersection). *For* S *in Equation* (2) *with* $x_n \neq \bar{x}_{(n)}$, *the fixed point* P *corresponding to the observation* $R(x_n, y_n)$ *is the point of intersection of the regression line for the set* S *and the regression line for* $S \setminus \{(x_n, y_n)\}$, *i.e., for* $\{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\}$.

Proof. Eliminating $x_n - \bar{x}_{(n-1)}$ between Formulas (28) and (30) yields

$$q = \bar{y}_{(n-1)} + \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)\left(y_i - \bar{y}_{(n-1)}\right)}{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)^2}\left(p - \bar{x}_{(n-1)}\right).$$

Thus, the fixed point $P(p,q)$ is on the least squares line

$$y = \bar{y}_{(n-1)} + \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)\left(y_i - \bar{y}_{(n-1)}\right)}{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)^2}\left(x - \bar{x}_{(n-1)}\right) \tag{31}$$

for the first $n-1$ observations of S. By Theorem 1, P is also on the regression line for S. ■

If (x_n, y_n) is on the line in Equation (31), then the two regression lines coincide.

If $x_n = \bar{x}_{(n)}$, then the fixed point is at infinity by Theorem 1. Also, using Identities (23) and (24), the slope of the regression line for S is

$$\begin{aligned}
&\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}_{(n)}\right)\left(y_i - \bar{y}_{(n)}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}_{(n)}\right)^2} \\
&= \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)\left(y_i - \bar{y}_{(n-1)}\right) + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)\left(y_n - \bar{y}_{(n-1)}\right)}{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)^2 + \frac{n-1}{n}\left(x_n - \bar{x}_{(n-1)}\right)^2} \\
&= \frac{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)\left(y_i - \bar{y}_{(n-1)}\right)}{\sum_{i=1}^{n-1}\left(x_i - \bar{x}_{(n-1)}\right)^2}.
\end{aligned}$$

which is the slope of the regression line for $S \setminus \{(x_n, y_n)\}$. If in addition, $y_n = \bar{y}_{(n)}$, then the two lines coincide. If $y_n \neq \bar{y}_{(n)}$, then the two lines have different points of means and are parallel, so that they "meet" at the fixed point at infinity.

The line in Equation (31) can be expressed with only basic sums as

$$y=\frac{\sum_{i=1}^{n-1}y_i}{n-1}+\frac{(n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i}{(n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2}\left(x-\frac{\sum_{i=1}^{n-1}x_i}{n-1}\right). \quad (32)$$

4. Additional Replacement Points

The question addressed in this section is:

Are there any other points in the plane, besides those described in Theorem 1, *that together with* $S\setminus\{(x_n,y_n)\}$, *i.e.,* $\{(x_i,y_i)\mid i=1,2,\cdots,n-1\}$, *produce regression lines containing the fixed point* $P(p,q)$ *corresponding to* (x_n,y_n)?

This question is answered in the affirmative by Theorems 3 and 4, where all such points are found.

Theorem 3 (Points on the Regression Line for $S\setminus\{(x_n,y_n)\}$). *Augmenting the set* $S\setminus\{(x_n,y_n)\}$ *with any one point of the line in Equation* (32), *which is the regression line on that set, yields a line containing the fixed point* $P(p,q)$ *that corresponds to the observation* (x_n,y_n).

Proof. This is immediate from the fact that adding a new point that is on a regression line does not alter the identity of the regression line and that P is on the line in Equation (32) by Theorem 2. ■

Theorem 3 shows that the commonplace diagnostic method of deleting a point to determine its effect is included in the technique of altering the point as in Theorem 1. Replacing the point (x_n,y_n) with the point at the intersection of the regression line for $S\setminus\{(x_n,y_n)\}$ and the line $x=x_n$ produces the regression line for the set $S\setminus\{(x_n,y_n)\}$. The nth point has been "neutralized," because it no longer has any role in determining the identity of the regression line, which has become the line on the other $n-1$ observations. This uses the least-squares estimate for the observation at $x=x_n$ that is based upon the other data, which is a familiar method for missing data [7] ([8], pp. 32-34) and for outliers [9].

Theorem 4 (All nth Points that Create the Same Fixed Point). *Augmenting the set* $S\setminus\{(x_n,y_n)\}=\{(x_i,y_i)\mid i=1,2,\cdots,n-1\}$ *with any single point on the vertical line in Equation* (16) *or on the regression line in Equation* (32) *determines a regression line containing the point* $P(p,q)$, *which is the fixed point that is determined by the observation* (x_n,y_n). *Those are the only single points in the plane that determine a regression line containing* P *by augmenting the set* $S\setminus\{(x_n,y_n)\}$.

Proof. The *proof* proceeds by replacing the chosen nth data point (x_n,y_n) of S with the point $A(s,t)$ to form the set $S\setminus\{(x_n,y_n)\}\cup\{(s,t)\}$. Substituting the coordinates of P from (26) into the regression line for that set yields an equation for s and t, which are the desired coordinates. That equation reduces to only two factors. One yields the vertical line in Equation (16). and the other represents the line in Equation (32). Because the *proof* consists principally of algebraic manipulations, it is placed in Appendix. ■

5. Geometry of Fixed Points

It is shown in this section that fixed points are centers about which diagnostic techniques are referenced.

From Equation (15), the regression line for set S is

$$y=\sum_{i=1}^{n}\left(\frac{1}{n}+\frac{\left(x_i-\overline{x}_{(n)}\right)\left(x-\overline{x}_{(n)}\right)}{\sum_{k=1}^{n}\left(k-\overline{x}_{(n)}\right)^2}\right)y_i .$$

The coefficients measure the impacts of the x-coordinates on the fitted values. The fitted value at the j th observation's x-value x_j is

$$y\left(x_j\right)=\sum_{i=1}^{n}\left(\frac{1}{n}+\frac{\left(x_i-\overline{x}_{(n)}\right)\left(x_j-\overline{x}_{(n)}\right)}{\sum_{k=1}^{n}\left(k-\overline{x}_{(n)}\right)^2}\right)y_i . \quad (33)$$

The parentheses in Equation (33) contain the terms of the hat matrix, which are designated h_{ij} [4] ([10], pp. 90-92). They show that the farther an observation's first coordinate is from the mean, the larger its impact on a fitted value. This effect is called leverage. Most often, the goal is to determine the impact that each observation has on the fitted value for the observation's x-value, so the diagonal elements h_{ii} of the hat matrix are especially important. They are written h_i.

Consider the same n th observation as in the previous sections. Its influence on $y\left(x_n\right)$ is $h_n y_n$. An interpretation is that changing y_n by one unit changes the fitted value $y\left(x_n\right)$ by h_n units. The set $\left\{h_i \mid i=1,2,\cdots,n\right\}$ has desirable properties. For example, $\sum_{i=1}^{n} h_i=2$, which supplies a finite scale for these positive terms. Two approaches to diagnosing whether an observation's x-value has high leverage is to compute the n values in the set $\left\{h_i\right\}$. The large values would indicate that their observations have high leverage. There are rules of thumb, as well [4] ([10], p. 91). One is that any observation with $h_i>0.5$ is problematical.

Sometimes, an observation is deleted in order to determine the observation's impact on some feature of the fit. From Theorem 3, deleting an observation yields the line for the remaining observations and that is the same line as the one obtained by adding a point on the line for the remaining observations. There are other properties and statistics that may be of interest in a study, such as the correlation coefficient. Those might be altered by the addition of a point. If the identity of the line or its slope are the attributes under study, an alternative to deletion is moving the point to the fit of the remaining observations [9].

Theorem 5 fleshes out the geometric picture.

Theorem 5 (Three Collinear Points). *The point of means* $\left(\overline{x}_{(n-1)},\overline{y}_{(n-1)}\right)$ *for* $\left\{\left(x_i,y_i\right) \mid i=1,2,\cdots,n-1\right\}$, *the point of means*

$$\left(\left(\sum_{i=1}^{n-1}x_i+s\right)\Big/n,\left(\sum_{i=1}^{n-1}y_i+t\right)\Big/n\right)=\left(\left((n-1)\overline{x}_{(n-1)}+s\right)\Big/n,\left((n-1)\overline{y}_{(n-1)}+t\right)\Big/n\right)$$

for $\left\{\left(x_i,y_i\right) \mid i=1,2,\cdots,n-1\right\}\cup\left\{(s,t)\right\}$, *and the point* (s,t) *are collinear or else coincide.*

Proof. Assuming the three points do not coincide, the line containing the two points of means is

$$y = \bar{y}_{(n-1)} + \frac{\left((n-1)\bar{y}_{(n-1)} + t\right)/n - \bar{y}_{(n-1)}}{\left((n-1)\bar{x}_{(n-1)} + s\right)/n - \bar{x}_{(n-1)}}\left(x - \bar{x}_{(n-1)}\right),$$

i.e.,

$$y = \bar{y}_{(n-1)} + \frac{t - \bar{y}_{(n-1)}}{s - \bar{x}_{(n-1)}}\left(x - \bar{x}_{(n-1)}\right). \tag{34}$$

The point (s,t) satisfies Equation (34). ■

In Theorem 5, if $s \neq \bar{x}_{(n-1)}$ and $t = \bar{y}_{(n-1)}$, then the line is horizontal. If $s = \bar{x}_{(n-1)}$ and $t \neq \bar{y}_{(n-1)}$, then the line is vertical. If $s = \bar{x}_{(n-1)}$ and $t = \bar{y}_{(n-1)}$, the three points coincide.

To summarize, by Theorem 1, the fixed point P, which corresponds to the chosen datum $R(x_n, y_n)$ for observations S, is a "center of rotation" with "spokes" being all the regression lines on the sets $S \setminus \{(x_n, y_n)\} \cup T(x_n, t)$. See Figure 1 and Figure 2. Theorem 2 says that the regression line for $S \setminus \{(x_n, y_n)\}$ is one of these spokes. The point of intersection V of $x = x_n$ and that regression line can be used to form the set $S \setminus \{(x_n, y_n)\} \cup V$, whose regression line is the same line as the regression line for $S \setminus \{(x_n, y_n)\}$ by Theorem 3. Point V is an example of a point that is generically labeled T. There are two points of means on this regression line. One is $N\left(\bar{x}_{(n-1)}, \bar{y}_{(n-1)}\right)$, and the other one, W, is for the set $S \setminus \{(x_n, y_n)\} \cup V$ and is on the line $x = \bar{x}_{(n)}$. The point of means for the set $S \setminus \{(x_n, y_n)\} \cup T$ is $M\left(\bar{x}_{(n-1)}, \left((n-1)\bar{y}_{(n-1)} + t\right)/n\right)$.

Figure 3 contains the regression line for $\{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\}$, which is the line with the larger positive slope in the graph, the points $R(x_n, y_n)$, N, W, and V, and fixed point P corresponding to R. The righthand (dashed) vertical line $x = x_n$ contains $R, T(x_n, t)$, and V. The lefthand (dashed) vertical line is $x = \bar{x}_{(n)}$, which contains M and W. Points M, N, and T are collinear by Theorem 5. The line containing P and M is the regression line for $\{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\} \cup T$.

The "motion" in the system is obtained by varying up and down the position of T. This moves the regression line's intersection with $x = x_n$ proportionally with h_n through the method of least-squares fitting. As T is moved, the point of means M of $\{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\} \cup T$ moves and is on the regression line that is being created, so that the regression line contains P and M. This can be performed with linkages; as T moves, the line containing M, N, and T rotates about N as the center and, thus, moves M vertically on $x = \bar{x}_{(n)}$. The line containing M and P rotates about fixed point P as the center and is the regression line.

Figure 4 illustrates a mechanical device that is a physical embodiment of the effect that one datum has on the location of the regression line. The datum's fixed point is a center. The blue slider on the right is moved by the operator and the green slider on the left follows, being controlled by the rods. The yellow items are centers of rotation; the red and the tan rods are attached there like hands of a clock. The goal of the device is to move the tan rod in a manner so that its locations reflect least-squares fitting of a regression line as the position of an observation, *i.e.*, the blue slider, is altered vertically by the operator.

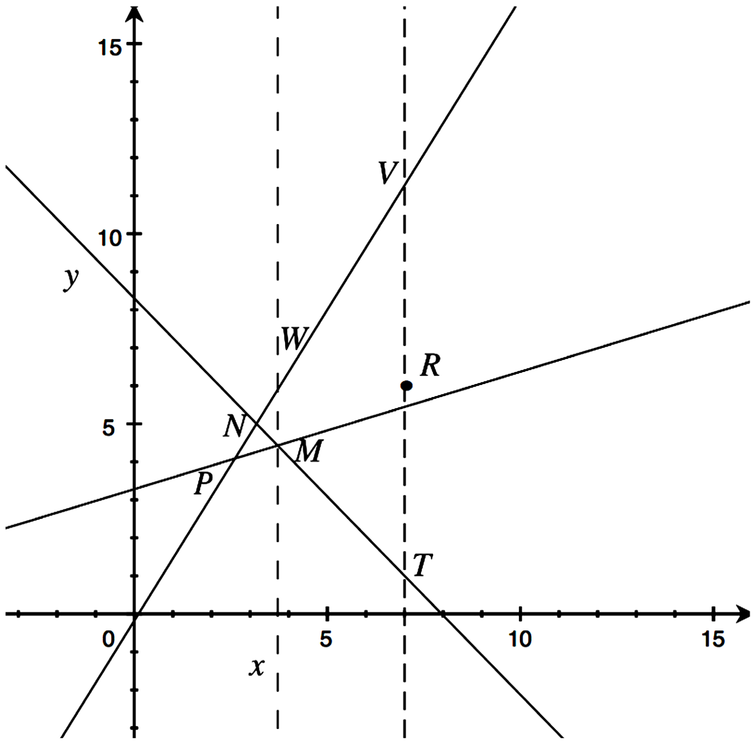

Figure 3. The observation $R(x_n, y_n)$ is replaced by $T(x_n, t)$.

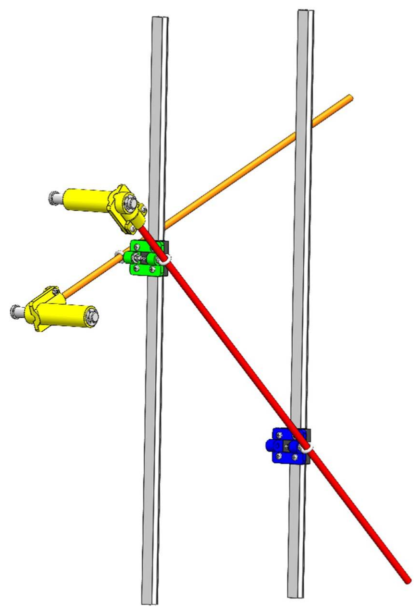

Figure 4. Mechanism to "move" the regression line (tan rod) by "moving" a data point (blue slider). See **Figure 3**, where the configuration of the points and lines is the same. In this figure, some points and lines are omitted. In particular, point R and the line containing points P, N, W, and V are omitted, because they are not directly involved in the mechanism.

When the mechanism is being run, the blue slider (T) is moved up and down continuously in a controlled manner on the righthand vertical rectangular grey track ($x = x_n$). The red rod (line containing M, N, and T), which passes through swiveling eyebolts, rotates about the upper yellow center (N), guided by and dragging along the green slider (M) on the lefthand vertical rectangular grey track ($x = \overline{x}_{(n)}$). As the green slider (M) is thusly moved up and down indirectly by the movement of the blue slider (T), the tan rod (line containing P and M) moves freely through a swiveling eyebolt, rotating about the lower yellow center (fixed point P) and is aligned with the regression line (for $\{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\} \cup T$).

6. Numerical Example

Consider the set of seven artificial bivariate observations

$$(1,3),\ (2,2),\ (3,3),\ (3,5),\ (5,8),\ (5,9),\ (7,6). \tag{35}$$

The selected point is $R(7,6)$. These data were used to create **Figure 1**. From Formula (26), the fixed point corresponding to R is

$$P(p,q) = \left(\frac{19}{6} - \frac{(6)(73)-(19)^2}{6((6)(7)-19)}, \frac{30}{6} - \frac{(6)(116)-(19)(30)}{(6)(7)-19} \right) = \left(\frac{60}{23}, \frac{94}{23} \right).$$

Replacing $y_n = y_7 = 6$ with 14, 124/11, 9, 6, 1, −2, and −4 in turn, the regression lines in **Figure 1** are obtained. As the replacement value decreases, the slopes of the corresponding lines in that graph decrease. The value 124/11 is the y-coordinate of the intersection of the regression line in Equation (32) on the first six observations and the line in Equation (16), $x = x_n = x_7 = 7$. From Theorems 2 and 3, the replacement value 124/11 yields the regression line on the first six observations, which is one of the regression lines that are found with the construction in Theorem 1 and contains the fixed point. In **Figure 1**, it is the line with the second largest slope. Among the regression lines displayed in **Figure 1**, the regression line for the original set of all the Observations (35) is the central line with slope that is approximately one.

The elements h_i of the hat matrix in the same order as the Observations (35) are 0.433, 0.259, 0.326, 0.326, 0.416, 0.416, 0.568.

Only h_7 is greater than 0.5, indicating that it has high leverage. This high leverage is in the context of the best-fit line. The z-score of $x = 7$ among the x-values is just 1.60, so that the point $(7,6)$ would not be considered an outlier as a member of the univariate set of x-values, but it appears to be impactful for the linear fitting.

7. Concluding Comments

The fixed points are far more fundamental and influential than previously realized. Literally, they are at the centers of regression diagnostic tests that involve altering or deleting an observation. The mechanical device in **Figure 4** supplies clear insight into the movements of points in the processes of alteration and deletion of an observation and shows the scales of related measurements. This can be useful for visual learners. Instead of just line graphs, the movements can be observed or imagined with the mechanical device.

The two main facts that make the construction in **Figure 3** and the mechanical device in **Figure 4** possible are Theorem 1, giving the existence of the fixed point P, and Theorem 5, giving collinearity of N, M, and T. The figures show that there are two centers, which are the yellow items in **Figure 4**. Careful examination of the formulas, shows that they can be written so that every occurrence of x-and y-values has the appropriate coordinate of $N\left(\overline{x}_{(n-1)}, \overline{y}_{(n-1)}\right)$ subtracted from it. Thus, N is a center and a natural choice for the origin, if a translation of the coordinates were contemplated. The other center is the fixed point P.

Often, the experimental design mandates that the x-values are pre-selected, while most often considered to be without error as well, and the y-values are measurements regarded as realizations of a random variable. That is one reason that the direction of prediction is perpendicular to the x-axis and that direction is of interest here. In those cases, the leverages and the *x*-coordinates of the pivot points can be computed before the data are taken.

The distinction between an observation having a potentially large or an actually large impact on the fitted line should be emphasized ([3], pp. 384-420) [4] ([10], pp. 90-92). A *high leverage point* has a substantial potential to greatly alter the position of the best-fit line. A high leverage point has a sizeable lever arm, *i.e.*, it is an inordinately large horizontal distance from the point of means within the framework of linear fitting. This is shown by the effect of each point on the slope as expressed in Formula (14) for example, where points with the largest numerators in the fractions have the highest leverage. Often, leverage is assessed using the set $\{h_i \mid i = 1, 2, \cdots, n\}$ of elements of the hat matrix, as seen in Section 5. However, high leverage points may not have a large impact on the fitted line, if their y-values place them near the line that would be determined by the remainder of the observations.

An *influential point* actually has a strong effect on the equation of the regression line for a given data set. One way to detect a single influential data point is to move it as is done above on a vertical line or to delete it and evaluate the sizes of the changes in the line. It is natural to examine the effect of varying the y-coordinate of a data point, because that action reflects procedures used to examine the influence of a point on a linear fit. Also, the impact of the deletion of a point can be assessed with the movement of the data point vertically to the regression line for the remainder of the observations, so that the slope snaps from its original value to the value it would have without the chosen point. All of these changes in the y-coordinate produce changes that have a fixed point as their center.

Any family of lines, whose coefficients can be parameterized as in Equation (19) has a fixed point. For example, all bivariate observations whose systems of fitted lines have coefficients that are linear functions of the observations' y-values have fixed points. A set of observations that is fitted with a simple least-squares regression line is an example, as shown with the second *proof* of Theorem 1. More generally, all lines

$$y = \left(q - pm(r)\right) + m(r)x,$$

where r is a real parameter that may be multidimensional, contain the point (p, q). Since models in general linear modeling and multiple regression have a linear dependence on the y -values, there are fixed lines and planes of one less dimension than the number of parameters or coefficients in the model [11].

In [11], the author considers a different problem in which the chosen observation occurs with a multiplicity and the regression line pivots about a point as the multiplicity changes. This is closely related to moving a single point [1]. In [11], the multiplicity is discrete, but here the selected observation can be moved continuously.

Acknowledgements

The author is deeply grateful to Hilary L. Farnsworth, Managing Partner, Instill Tech, for the care taken with **Figure 4**, which was created using a 3D modeling tool.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Lutzer, C.V. and Farnsworth, D.L. (2021) Pivot Points in Bivariate Linear Regression. *Open Journal of Statistics*, **11**, 393-399. https://doi.org/10.4236/ojs.2021.113023

[2] Montgomery, D.C., Runger, G.C. and Hubele, N.F. (2011) Engineering Statistics. 5th Edition, Wiley.

[3] Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005) Applied Linear Statistical Models. 5th Edition, McGraw-Hill/Irwin. https://doi.org/10.2307/1269508

[4] Hoaglin, D.C. (1988) Using Leverage and Influence to Introduce Regression Diagnostics. *The College Mathematics Journal*, **19**, 387-416. https://doi.org/10.1080/07468342.1988.11973146

[5] Chatterjee, S. and Hadi, A.S. (1986) Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, **1**, 379-393. https://doi.org/10.1214/ss/1177013622

[6] Chatterjee, S. and Hadi, A.S. (1988). Sensitivity Analysis in Linear Regression. Wiley. https://doi.org/10.1002/9780470316764

[7] Carpenter J.R. and Kenward, M.G. (2015) Development of Methods for the Analysis of Partially Observed Data and Critique of Ad Hoc Methods. In: Molenberghs, G., Fitzmaurice, G., Kenward, M.G., Tsiatis, A. and Verbeke, G., Eds., *Handbook of Missing Data Methodology*, Chapman & Hall/CRC Press, 23-46. https://doi.org/10.1201/b17622

[8] Little, R. and Rubin, D. (2019) Statistical Analysis with Missing Data. 3rd Edition, Wiley. https://doi.org/10.1002/9781119482260

[9] David L. Farnsworth, (2024) Analysis of Data Containing Outliers. *Journal of Prob-*

ability and Statistical Science, **22**, 99-105. https://doi.org/10.37119/jpss2024.v22i1.794

[10] Harrell, Jr. (2015) Regression Modeling Strategies with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. 2nd Edition, Springer. https://doi.org/10.1007/978-3-319-19425-7

[11] Lutzer, C.V. (2017) A Curious Feature of Regression. *The College Mathematics Journal*, **48**, 189-198. https://doi.org/10.4169/college.math.j.48.3.189

Appendix

Proof of Theorem 4

Replacing the nth data point (x_n, y_n) of S with $A(s,t)$, which is to be determined, the least-squares regression line for the set

$$S \setminus \{(x_n, y_n)\} \cup \{(s,t)\} = \{(x_i, y_i) \mid i = 1, 2, \cdots, n-1\} \cup \{(s,t)\}$$

is

$$y = \frac{\sum_{i=1}^{n-1} y_i + t}{n} + \frac{n\left(\sum_{i=1}^{n-1} x_i y_i + st\right) - \left(\sum_{i=1}^{n-1} x_i + s\right)\left(\sum_{i=1}^{n-1} y_i + t\right)}{n\left(\sum_{i=1}^{n-1} x_i^2 + s^2\right) - \left(\sum_{i=1}^{n-1} x_i + s\right)^2}\left(x - \frac{\sum_{i=1}^{n-1} x_i + s}{n}\right)$$

(compare with the regression line in Equation (10)). In order to find the possible coordinates for point A, the coordinates of P as expressed in Formula (26) are substituted for (x, y) in the last equation, giving

$$\begin{aligned}
&\frac{\sum_{i=1}^{n-1} y_i}{n-1} - \frac{(n-1)\sum_{i=1}^{n-1} x_i y_i - \sum_{i=1}^{n-1} x_i \sum_{i=1}^{n-1} y_i}{(n-1)\left((n-1)x_n - \sum_{i=1}^{n-1} x_i\right)} - \frac{\sum_{i=1}^{n-1} y_i + t}{n} \\
&= \left(\frac{n\left(\sum_{i=1}^{n-1} x_i y_i + st\right) - \left(\sum_{i=1}^{n-1} x_i + s\right)\left(\sum_{i=1}^{n-1} y_i + t\right)}{n\left(\sum_{i=1}^{n-1} x_i^2 + s^2\right) - \left(\sum_{i=1}^{n-1} x_i + s\right)^2}\right) \\
&\times \left(\frac{\sum_{i=1}^{n-1} x_i}{n-1} - \frac{(n-1)\sum_{i=1}^{n-1} x_i^2 - \left(\sum_{i=1}^{n-1} x_i\right)^2}{(n-1)\left((n-1)x_n - \sum_{i=1}^{n-1} x_i\right)} - \frac{\sum_{i=1}^{n-1} x_i + s}{n}\right).
\end{aligned}$$

Then,

$$\begin{aligned}
&\frac{(n-1)x_n \sum_{i=1}^{n-1} y_i - \sum_{i=1}^{n-1} x_i \sum_{i=1}^{n-1} y_i - n\left((n-1)\sum_{i=1}^{n-1} x_i y_i - \sum_{i=1}^{n-1} x_i \sum_{i=1}^{n-1} y_i\right) - (n-1)^2 t x_n + (n-1)t\sum_{i=1}^{n-1} x_i}{n(n-1)\left((n-1)x_n - \sum_{i=1}^{n-1} x_i\right)} \\
&= \left(\frac{n\sum_{i=1}^{n-1} x_i y_i - \sum_{i=1}^{n-1} x_i \sum_{i=1}^{n-1} y_i + (n-1)st - s\sum_{i=1}^{n-1} y_i - t\sum_{i=1}^{n-1} x_i}{n\sum_{i=1}^{n-1} x_i^2 - \left(\sum_{i=1}^{n-1} x_i\right)^2 + (n-1)s^2 - 2s\sum_{i=1}^{n-1} x_i}\right) \\
&\times \left(\frac{(n-1)x_n \sum_{i=1}^{n-1} x_i - \left(\sum_{i=1}^{n-1} x_i\right)^2 - n\left((n-1)\sum_{i=1}^{n-1} x_i^2 - \left(\sum_{i=1}^{n-1} x_i\right)^2\right) - (n-1)^2 s x_n + (n-1)s\sum_{i=1}^{n-1} x_i}{n(n-1)\left((n-1)x_n - \sum_{i=1}^{n-1} x_i\right)}\right)
\end{aligned}$$

or

$$\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\sum_{i=1}^{n-1}y_i-\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)(n-1)t-n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)}{n(n-1)\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)}$$

$$=\left(\frac{\frac{n}{n-1}\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i+\frac{1}{n}\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)+\frac{(n-1)^2st-(n-1)s\sum_{i=1}^{n-1}y_i-(n-1)t\sum_{i=1}^{n-1}x_i}{n-1}}{\frac{n}{n-1}\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2+\frac{1}{n}\left(\sum_{i=1}^{n-1}x_i\right)^2\right)+\frac{(n-1)^2s^2-2(n-1)s\sum_{i=1}^{n-1}x_i}{n-1}}\right)$$

$$\times\left(\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\sum_{i=1}^{n-1}x_i-\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)(n-1)s-n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)}{n(n-1)\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)}\right)$$

or

$$\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\sum_{i=1}^{n-1}y_i-\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)(n-1)t-n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)}{n(n-1)\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)}$$

$$=\left(\frac{\frac{1}{n-1}\left((n-1)^2st-(n-1)s\sum_{i=1}^{n-1}y_i-(n-1)t\sum_{i=1}^{n-1}x_i+\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\right)}{\frac{1}{n-1}\left((n-1)^2s^2-2(n-1)s\sum_{i=1}^{n-1}x_i+\left(\sum_{i=1}^{n-1}x_i\right)^2+n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\right)}\right)$$

$$\times\left(\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\sum_{i=1}^{n-1}x_i-\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)(n-1)s-n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)}{n(n-1)\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)}\right).$$

This becomes

$$\left(\frac{-1}{n(n-1)}\right)\left(\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)}{(n-1)x_n-\sum_{i=1}^{n-1}x_i}\right)$$

$$=\left(\frac{\frac{1}{n-1}\left(\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\right)}{\frac{1}{n-1}\left(\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)^2+n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\right)}\right)$$

$$\times\left(\frac{-1}{n(n-1)}\right)\left(\frac{\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)}{(n-1)x_n-\sum_{i=1}^{n-1}x_i}\right)$$

or

$$\left(\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\right)$$
$$\times\left(\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)^2+n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\right)$$
$$=\left(\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\right)$$
$$\times\left(\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)+n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\right).$$

Multiplying yields

$$\begin{aligned}
&\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)^2\\
&+\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)^2 n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\\
&+\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\\
&+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\\
&=\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\\
&+\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\\
&+\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\\
&+n\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)n\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right).
\end{aligned}$$

Observing that the first terms on the two sides of the last equation are the same, as are the fourth terms the same, this becomes

$$\begin{aligned}
&\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)^2\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\\
&+\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\\
&=\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\\
&+\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)
\end{aligned}$$

subsequent to dividing by n. Re-arranging yields

$$\begin{aligned}
&\left[\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)-\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\right]\\
&\times\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\\
&=\left[\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)-\left((n-1)x_n-\sum_{i=1}^{n-1}x_i\right)\right]\\
&\times\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right),
\end{aligned}$$

and, by re-arranging further and dividing by $n-1\neq 0$,

$$\begin{aligned}
&[x_n-s]\left[\left((n-1)t-\sum_{i=1}^{n-1}y_i\right)\left((n-1)\sum_{i=1}^{n-1}x_i^2-\left(\sum_{i=1}^{n-1}x_i\right)^2\right)\right.\\
&\left.-\left((n-1)s-\sum_{i=1}^{n-1}x_i\right)\left((n-1)\sum_{i=1}^{n-1}x_iy_i-\sum_{i=1}^{n-1}x_i\sum_{i=1}^{n-1}y_i\right)\right]=0.
\end{aligned}$$

The first factor in square brackets gives the vertical line in Equation (16); the second factor gives the regression line in Equation (32); and there are no other solutions.