Applied Mathematical Sciences, Vol. 6, 2012, no. 57, 2821 - 2832

Corollary to a Theorem of Oblique Linear Regression

Brian J. McCartin

Applied Mathematics, Kettering University
1700 University Avenue, Flint, MI 48504-6214 USA
bmccarti@kettering.edu

Abstract

The Galton-Pearson-McCartin Theorem [5, p. 2899] states that the best and worse lines of oblique regression contain conjugate diameters of the concentration and inertia ellipses. The First Theorem of Apollonius [9, p. 100] states that the sum of squares of any pair of conjugate semi-diameters of an ellipse is a constant. These theorems are herein combined to yield a relation between the orthogonal sums of squares of the best and worse lines of oblique regression that is independent of the direction of oblique projection.

Mathematics Subject Classification: 62J05

Keywords: oblique linear regression, concentration ellipse, inertia ellipse, conjugate diameters, Galton-Pearson-McCartin theorem

1 Introduction

If, in fitting a linear relationship to a set of experimental (x, y) data (Figure 1), it is assumed that one ('independent') variable, say x, is known exactly while the other ('dependent') variable, say y, is subject to error then the line L (Figure 2) is chosen to minimize the total square vertical deviation Σd_y^2 and is known as the line of *regression of y on x*. Reversing the roles of x and y minimizes instead the total square horizontal deviation Σd_x^2, thereby yielding the line of *regression of x on y*. Either method will be termed *coordinate regression* [2].

This procedure may be generalized to the situation where both variables are subject to independent errors with zero means, albeit with equal variances. In this case, L is chosen to minimize the total square orthogonal deviation $\Sigma d_\perp^2$ (Figure 2). This will be referred to as *orthogonal regression* [7].

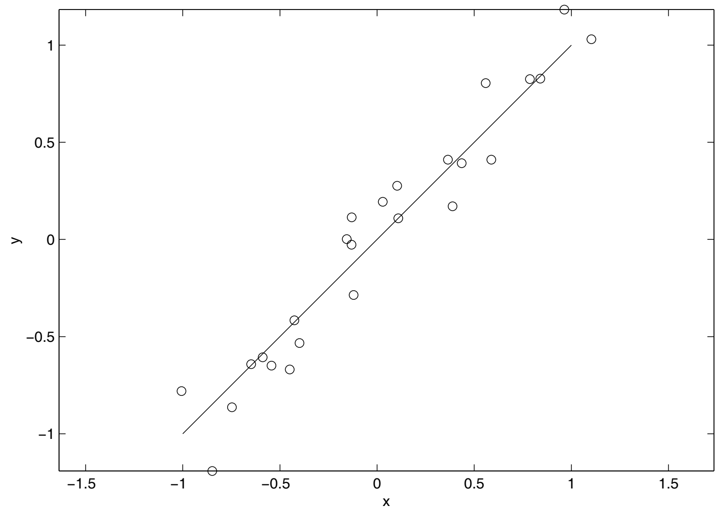

Figure 1: Fitting Experimental Data

Moreover, this approach may be further extended to embrace the case where the error variances are unequal. Defining σ_u^2 = variance of x-error, σ_v^2 = variance of y-error and $\lambda = \sigma_v^2/\sigma_u^2$, a weighted least squares procedure is employed whereby $\frac{1+m^2}{\lambda+m^2} \cdot \Sigma d_\perp^2$ is minimized in order to determine L. This will be referred to as *λ-regression* [3]. Finally, if the errors are correlated with p_{uv} = covariance of errors and $\mu = p_{uv}/\sigma_u^2$ then $\frac{1+m^2}{\lambda-2\mu m+m^2} \cdot \Sigma d_\perp^2$ is minimized in order to determine L. This will be referred to as *(λ, μ)-regression* [4].

The concept of *oblique linear regression* (Figure 3) was introduced in [5]. In this mode of approximation, we first select a direction specified by the slope s. We then obliquely project each data point in this direction onto the line L. Finally, L is then chosen so as to minimize the total square oblique deviation Σd_i^2. It was therein demonstrated that this permits a unified treatment of all of the previously described flavors of linear regression.

Galton [2] provided a geometric interpretation of coordinate regression in terms of the best-fitting *concentration ellipse* which has the same first moments and second moments about the centroid as the experimental data [1, pp. 283-285]. Pearson [7] extended this geometric characterization to orthogonal regression. McCartin [3] further extended this geometric characterization to λ-regression. McCartin [4] finally succeeded in extending this geometric characterization to the general case of (λ, μ)-regression. In a similar vein, a geometric characterization of oblique linear regression [5], the Galton-Pearson-McCartin (GPM) Theorem, was proffered which was subsequently shown to further generalize these Galton-Pearson-McCartin geometric characterizations of linear regression.

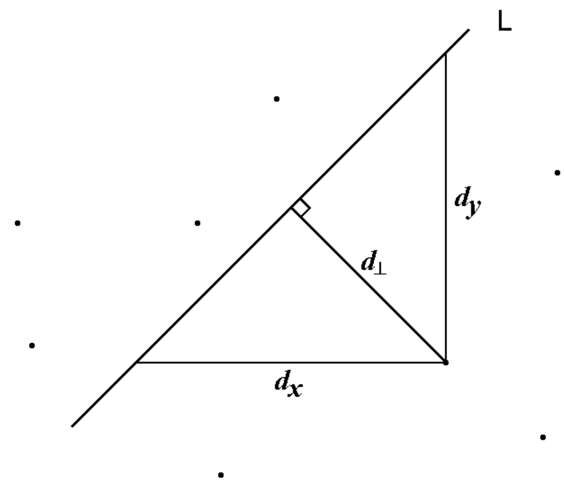

Figure 2: Coordinate and Orthogonal Regression

In the present paper, the First Theorem of Apollonius [9, p. 100], to the effect that the sum of squares of any pair of conjugate semi-diameters of an ellipse is a constant, is combined with the GPM Theorem to yield a corollary embodying a relationship between the orthogonal sums of squares of the best and worse lines of oblique regression that is independent of the direction of oblique projection. However, before proceeding any further, the elements of oblique linear regression are reviewed.

2 Oblique Linear Regression

Consider the experimentally 'observed' data $\{(x_i, y_i)\}_{i=1}^{n}$ where $x_i = X_i + u_i$ and $y_i = Y_i + v_i$. Here (X_i, Y_i) denote theoretically exact values with corresponding random errors (u_i, v_i). We shall assume that $E(u_i) = 0 = E(v_i)$, that successive observations are independent, and that $Var(u_i) = \sigma_u^2$, $Var(v_i) = \sigma_v^2$, $Cov(u_i, v_i) = p_{uv}$ irrespective of i.

Next, define the statistics of the sample data corresponding to the above population statistics. The mean values are given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i;\ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{1}$$

the sample variances are given by

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2;\ \sigma_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{2}$$

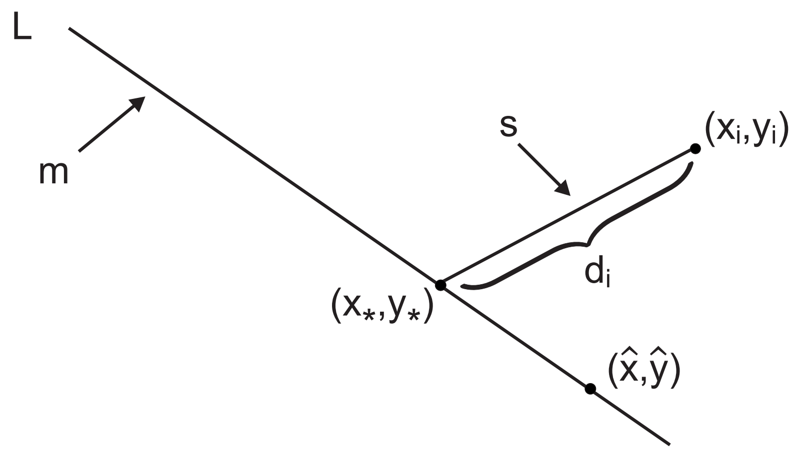

Figure 3: Oblique Projection

the sample covariance is given by

$$p_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y}), \tag{3}$$

and (assuming that $\sigma_x \cdot \sigma_y \neq 0$) the sample correlation coefficient is given by

$$r_{xy} = \frac{p_{xy}}{\sigma_x \cdot \sigma_y}. \tag{4}$$

Note that if $\sigma_x^2 = 0$ then the data lie along the vertical line $x = \bar{x}$ while if $\sigma_y^2 = 0$ then they lie along the horizontal line $y = \bar{y}$. Hence, we assume without loss of generality that $\sigma_x^2 \cdot \sigma_y^2 \neq 0$ so that r_{xy} is always well defined. Furthermore, by the Cauchy-Buniakovsky-Schwarz inequality,

$$p_{xy}^2 \leq \sigma_x^2 \cdot \sigma_y^2 \text{ (i.e. } -1 \leq r_{xy} \leq 1) \tag{5}$$

with equality if and only if $(y_i - \bar{y}) \propto (x_i - \bar{x})$, in which case the data lie on the line

$$y - \bar{y} = \frac{p_{xy}}{\sigma_x^2}(x - \bar{x}) = \frac{\sigma_y^2}{p_{xy}}(x - \bar{x}), \tag{6}$$

since $p_{xy} \neq 0$ in this instance. Thus, we may also restrict $-1 < r_{xy} < 1$.

Turning our attention to Figure 3, the oblique regression line, L, corresponding to the specified slope s may be expressed as

$$y - \hat{y} = m(x - \hat{x}), \tag{7}$$

where $(\hat{x}, \hat{y})$ is any convenient point lying on the line. L is to be selected so as to minimize the total square oblique deviation

$$S := \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}[(x_i - x_*)^2 + (y_i - y_*)^2]. \tag{8}$$

Straightforward analytic geometry implies that we should instead minimize

$$S = \frac{1+s^2}{(m-s)^2} \cdot \sum_{i=1}^{n} [m(x_i - \hat{x}) - (y_i - \hat{y})]^2. \tag{9}$$

Before proceeding any further, we point out an important simplification of our optimization problem:

Lemma 1 *The oblique regression line, defined as the solution to the above optimization problem, always passes through the centroid, $(\bar{x}, \bar{y})$, of the data.*

Thus, Equation (7) for the oblique regression line may be replaced by

$$L : y - \bar{y} = m(x - \bar{x}) \tag{10}$$

which passes through the centroid.

As a result of this, we may reduce our optimization problem to choosing m to minimize

$$S(m; s) = \frac{1+s^2}{(m-s)^2} \cdot \sum_{i=1}^{n} [m(x_i - \bar{x}) - (y_i - \bar{y})]^2, \tag{11}$$

for a given s. Furthermore, we first recast the objective function as

$$S(m; s) = \frac{1+s^2}{(m-s)^2} \cdot n \cdot [m^2 \sigma_x^2 - 2m p_{xy} + \sigma_y^2]. \tag{12}$$

Differentiating Equation (12) with respect to m and setting $S'(m)$ to zero produces the extremizing slope:

Theorem 1 *The slope of the oblique regression line is given by*

$$m^*(s) = \frac{\sigma_y^2 - s \cdot p_{xy}}{p_{xy} - s \cdot \sigma_x^2}. \tag{13}$$

Thereby, Equations (10) and (13) taken together comprise a complete solution to the problem of oblique linear regression.

Observe that, from Equation (11), oblique linear regression may be interpreted as a weighted least squares procedure whereby L is chosen to minimize $\frac{(1+m^2)\cdot(1+s^2)}{(m-s)^2} \cdot \Sigma d_\perp^2$. The symmetry of this objective function with respect to m and s leads directly to:

Corollary 1 (Reciprocity Principle) *If m is the slope of the oblique regression line in the direction s then s is the slope of the oblique regression line in the direction m.*

3 Special Cases of Oblique Linear Regression

We next consider some important special cases of oblique linear regression [5].

3.1 Coordinate Regression

We commence with the case of coordinate regression [2]. Setting $s = 0$ in Equation (13) produces regression of x on y:

$$m_x := \frac{\sigma_y^2}{p_{xy}}. \tag{14}$$

Likewise, letting $s \to \infty$ in Equation (13) produces regression of y on x:

$$m_y := \frac{p_{xy}}{\sigma_x^2}. \tag{15}$$

Unsurprisingly, we see that the coordinate regression lines are special cases of oblique linear regression.

3.2 Orthogonal Regression

We now consider the more substantial case of orthogonal regression [7]. Setting $s = -\frac{1}{m}$ in Equation (13) produces

$$p_{xy} \cdot m^2 - (\sigma_y^2 - \sigma_x^2) \cdot m - p_{xy} = 0, \tag{16}$$

whose roots provide both the best orthogonal regression line

$$m_\perp = \frac{(\sigma_y^2 - \sigma_x^2) + \sqrt{(\sigma_y^2 - \sigma_x^2)^2 + 4p_{xy}^2}}{2p_{xy}}, \tag{17}$$

as well as the worst orthogonal regression line

$$s_\perp = \frac{(\sigma_y^2 - \sigma_x^2) - \sqrt{(\sigma_y^2 - \sigma_x^2)^2 + 4p_{xy}^2}}{2p_{xy}}. \tag{18}$$

Thus, orthogonal regression is also a special case of oblique linear regression.

3.3 λ-Regression

Turning now to the case of λ-regression [3], set $s = -\frac{\lambda}{m}$ in Equation (13):

$$p_{xy} \cdot m^2 - (\sigma_y^2 - \lambda\sigma_x^2) \cdot m - \lambda p_{xy} = 0, \tag{19}$$

whose roots provide both the best λ-regression line

$$m_\lambda = \frac{(\sigma_y^2 - \lambda\sigma_x^2) + \sqrt{(\sigma_y^2 - \lambda\sigma_x^2)^2 + 4\lambda p_{xy}^2}}{2p_{xy}}, \tag{20}$$

as well as the worst λ-regression line

$$s_\lambda = \frac{(\sigma_y^2 - \lambda\sigma_x^2) - \sqrt{(\sigma_y^2 - \lambda\sigma_x^2)^2 + 4\lambda p_{xy}^2}}{2p_{xy}}. \tag{21}$$

Thus, λ-regression is also a special case of oblique linear regression.

3.4 (λ, μ)-Regression

Proceeding to the case of (λ, μ)-regression [4], set $s = \frac{\mu m - \lambda}{m - \mu}$ in Equation (13):

$$(p_{xy} - \mu\sigma_x^2) \cdot m^2 - (\sigma_y^2 - \lambda\sigma_x^2) \cdot m - (\lambda p_{xy} - \mu\sigma_y^2) = 0, \tag{22}$$

whose roots provide both the best (λ, μ)-regression line

$$m_{\lambda,\mu} = \frac{(\sigma_y^2 - \lambda\sigma_x^2) + \sqrt{(\sigma_y^2 - \lambda\sigma_x^2)^2 + 4(p_{xy} - \mu\sigma_x^2)(\lambda p_{xy} - \mu\sigma_y^2)}}{2(p_{xy} - \mu\sigma_x^2)}, \tag{23}$$

as well as the worst (λ, μ)-regression line

$$s_{\lambda,\mu} = \frac{(\sigma_y^2 - \lambda\sigma_x^2) - \sqrt{(\sigma_y^2 - \lambda\sigma_x^2)^2 + 4(p_{xy} - \mu\sigma_x^2)(\lambda p_{xy} - \mu\sigma_y^2)}}{2(p_{xy} - \mu\sigma_x^2)}. \tag{24}$$

Thus, (λ, μ)-regression is also a special case of oblique linear regression.

4 Concentration and Inertia Ellipses

Define the *concentration ellipse* (Figure 4) via

$$\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\frac{p_{xy}}{\sigma_x\sigma_y} \cdot \frac{(x-\bar{x})}{\sigma_x} \cdot \frac{(y-\bar{y})}{\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} = 4(1 - \frac{p_{xy}^2}{\sigma_x^2\sigma_y^2}). \tag{25}$$

It has the same centroid and second moments about that centroid as does the data [1, pp. 283-285]. In this sense, it is the ellipse which is most representative of the data points without any a priori statistical assumptions concerning their origin.

Define also the *inertia ellipse* (Figure 5) via

$$\frac{(x-\bar{x})^2}{\sigma_x^2} - 2\frac{p_{xy}}{\sigma_x\sigma_y} \cdot \frac{(x-\bar{x})}{\sigma_x} \cdot \frac{(y-\bar{y})}{\sigma_y} + \frac{(y-\bar{y})^2}{\sigma_y^2} = \frac{1}{n \cdot \sigma_x^2\sigma_y^2}. \tag{26}$$

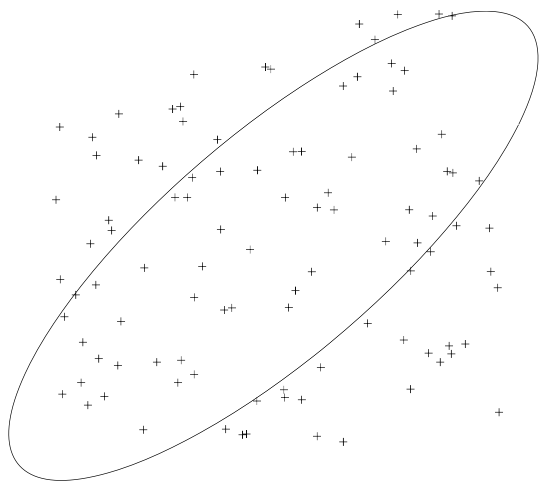

Figure 4: Concentration Ellipse

The reciprocal of the square of the distance from the centroid of any point on its periphery is equal to the moment of inertia of unit masses located at the data points with respect to the line joining the centroid to that peripheral point [1, pp. 275-276]. In turn, this moment of inertia is given by Equation (11) with $s = -\frac{1}{m}$, viz.

$$S_{\perp}(m) := S\left(m; -\frac{1}{m}\right) = \frac{1}{(1+m^2)} \cdot \sum_{i=1}^{n} [m(x_i - \bar{x}) - (y_i - \bar{y})]^2, \tag{27}$$

which is seen to be the sum of squares of orthogonal distances of the data points from the line with slope m passing through their centroid. Importantly, the concentration and inertia ellipses are clearly homothetic with one another with center of similitude located at the centroid.

In order to describe the GPM Theorem and its corollary in the next section, we will require the concept of conjugate diameters of an ellipse [8, p. 146]. A *diameter* of an ellipse is a line segment passing through the center connecting two antipodal points on its periphery. The *conjugate diameter* to a given diameter is that diameter which is parallel to the tangent to the ellipse at either peripheral point of the given diameter (see Figure 6, where the dashed line is the major axis).

If the ellipse is described by

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0 \ (h^2 < ab), \tag{28}$$

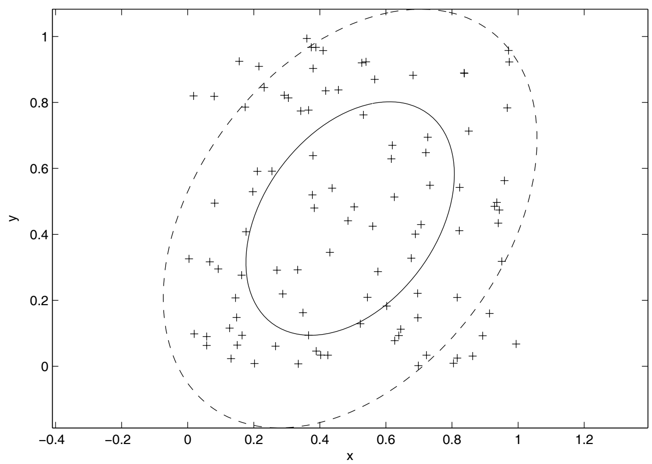

Figure 5: Concentration (dashed) and Inertia (solid) Ellipses

then, according to [8, p. 146], the slope of the conjugate diameter is

$$s = -\frac{a + h \cdot m}{h + b \cdot m}, \tag{29}$$

where m is the slope of the given diameter. Consequently,

$$m = -\frac{a + h \cdot s}{h + b \cdot s}, \tag{30}$$

thus establishing that conjugacy is a symmetric relation. As a result, these conjugacy conditions may be rewritten symmetrically as

$$b\ m \cdot s + h\ (m + s) + a = 0. \tag{31}$$

The final tool that is required is the following classical result [9, p. 100].

Theorem 2 (First Theorem of Apollonius) *The sum of squares of two conjugate semi-diameters is constant and equal to the sum of squares of the principal semi-axes.*

A matrix analytic proof of this important theorem is to be found in [6].

5 The GPM Theorem and Its Corollary

As established previously, all of the familiar modes of linear regression are subsumed under the umbrella of oblique linear regression. We now consider the

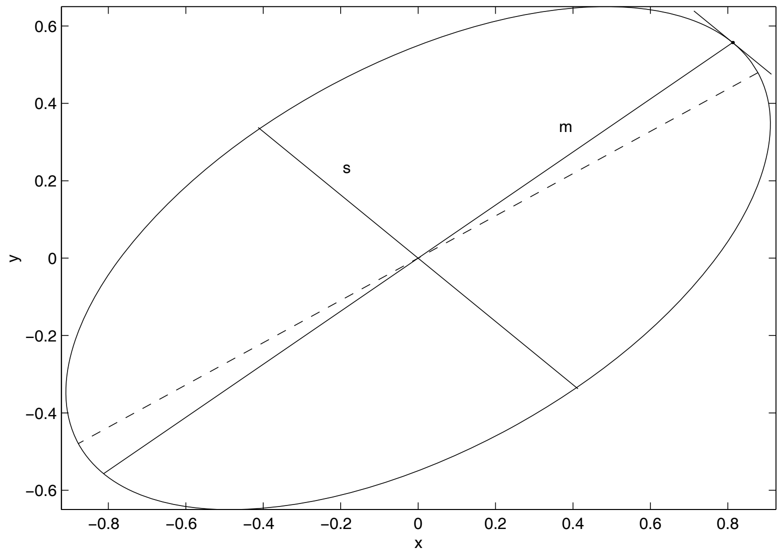

Figure 6: Conjugate Diameters

generalization of the Galton-Pearson-McCartin geometric characterizations of linear regression [2, 7, 3, 4] to oblique linear regression [5] and its immediate corollary.

For the concentration ellipse, Equation (25), and inertia ellipse, Equation (26), we may take

$$a = \sigma_y^2,\ h = -p_{xy},\ b = \sigma_x^2, \tag{32}$$

so that

$$s = -\frac{\sigma_y^2 - p_{xy} \cdot m}{-p_{xy} + \sigma_x^2 \cdot m};\ m = -\frac{\sigma_y^2 - p_{xy} \cdot s}{-p_{xy} + \sigma_x^2 \cdot s}, \tag{33}$$

which, by Equation (31), implies that s and m satisfy the symmetric conjugacy condition

$$\sigma_x^2\ m \cdot s - p_{xy}\ (m + s) + \sigma_y^2 = 0. \tag{34}$$

Direct comparison of Equations (33) and (13) demonstrates that the best and worst slopes of oblique regression satisfy Equation (34) and are thus conjugate to one another. Hence, we arrive at the unifying GPM Theorem [5]:

Theorem 3 (Galton-Pearson-McCartin Theorem) *The best and worst lines of oblique regression contain conjugate diameters of the concentration and inertia ellipses whose slopes satisfy the conjugacy relation Equation (34).*

Note that $s = \frac{\mu m - \lambda}{m - \mu}$ may be rearranged to reproduce the geometric characterization for (λ, μ)-regression of McCartin [4]: $(m - \mu) \cdot (s - \mu) = \mu^2 - \lambda$.

Furthermore, subsequently setting $\mu = 0$ reproduces the geometric characterization for λ-regression of McCartin [3]: $m \cdot s = -\lambda$. Now, setting $\lambda = 1$ reproduces Pearson's geometric characterization for orthogonal regression [7]: it is the major axis of the ellipse. Instead, setting $\lambda = 0, \infty$ reproduces Galton's geometric characterization for coordinate regression [2]: they are the lines connecting the points of horizontal/vertical tangencies, respectively.

Combining the GPM Theorem with the First Theorem of Apollonius leads immediately to the principal new result of the present paper:

Corollary 2 (GPM Corollary) *The orthogonal sums of squares of the best oblique regression line, $S_\perp(m)$, and the worse oblique regression line, $S_\perp(s)$, satisfy the relation:*

$$\left[\frac{1}{n}S_\perp(m)\right]^{-1} + \left[\frac{1}{n}S_\perp(s)\right]^{-1} = \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^2 - p_{xy}^2}, \tag{35}$$

independent of the direction of oblique projection s.

<u>Proof:</u> By the definition of the inertia ellipse, the sum of squares of a pair of conjugate semi-diameters is equal to the sum of the reciprocals of the corresponding moments of inertia given by Equation (27). By the First Theorem of Apollonius, this sum of reciprocals is constant and equal to that for the principal semi-axes. As the major and minor axes correspond to the best and worst orthogonal regression lines, respectively, the moment of inertia with respect to the major axis is [3, Equation (28)]

$$S_\perp(m_\perp) = \frac{n}{2} \cdot \left[(\sigma_y^2 + \sigma_x^2) - \sqrt{(\sigma_y^2 - \sigma_x^2)^2 + 4p_x y^2}\right],$$

while that with respect to the minor axis is [3, Equation (29)]

$$S_\perp(s_\perp) = \frac{n}{2} \cdot \left[(\sigma_y^2 + \sigma_x^2) + \sqrt{(\sigma_y^2 - \sigma_x^2)^2 + 4p_x y^2}\right].$$

Thus, reciprocating and adding these last two equations yields the desired relation:

$$\left[\frac{1}{n}S_\perp(m)\right]^{-1} + \left[\frac{1}{n}S_\perp(s)\right]^{-1} = \frac{\sigma_x^2 + \sigma_y^2}{\sigma_x^2\sigma_y^2 - p_{xy}^2}.$$

□

6 Conclusion

In the foregoing, the concept of oblique linear regression [5] was reviewed with emphasis placed upon its ability to provide a unified treatment of the various modes of linear regression: coordinate regression [2], orthogonal regression [7], λ-regression [3] and (λ, μ)-regression [4]. This development culminated in the Galton-Pearson-McCartin Theorem [5] and the opportunity to present and prove an interesting corollary using the First Theorem of Apollonius has been exploited.

Acknowledgement

The author expresses his sincere gratitude to Mrs. Barbara A. McCartin for her dedicated assistance in the production of this paper.

References

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, NJ, 1946.

[2] F. Galton, Family Likeness in Stature (with Appendix by J. D. H. Dickson), *Proc. Roy. Soc. London* 40 (1886), 42-73.

[3] B. J. McCartin, A Geometric Characterization of Linear Regression, *Statistics* 37(2) (2003), 101-117.

[4] B. J. McCartin, The Geometry of Linear Regression with Correlated Errors, *Statistics*, 39 (1) (2005), 1-11.

[5] B. J. McCartin, Oblique Linear Least Squares Approximation, *Applied Mathematical Sciences*, Vol. 4, No. 58, 2010, pp. 2891-2904.

[6] B. J. McCartin, A Matrix Analytic Approach to Conjugate Diameters of an Ellipse, *Applied Mathematical Sciences*, Vol. 6, 2012, to appear.

[7] K. Pearson, On Lines and Planes of Closest Fit to Systems of Points in Space, *Phil. Mag.* 2 (1901), 559-572.

[8] G. Salmon, *A Treatise on Conic Sections*, Sixth Edition, Chelsea, New York, NY, 1954.

[9] C. Zwikker, *The Advanced Geometry of Plane Curves and Their Applications*, Dover, Mineola, NY, 1963.

Received: January 2012