Technical Note: Regression Analysis in Adult Age Estimation

R.G. AYKROYD,[1] D. LUCY,[2]* A.M. POLLARD,[2] AND T. SOLHEIM[3]
[1]*Department of Statistics, School of Mathematics, University of Leeds, Leeds, West Yorkshire, United Kingdom LS2 9JT*
[2]*Department of Archaeological Sciences, University of Bradford, Bradford, West Yorkshire, United Kingdom BD7 1DP*
[3]*Department of Oral Pathology and Section for Forensic Odontology, University of Oslo, Blindern, 0317 Oslo, Norway*

KEY WORDS classical calibration; inverse calibration; regression; age estimation

ABSTRACT Accurate estimation of human adult age has always been a problem for anthropologists, archaeologists and forensic scientists. The main factor contributing to the difficulties is the high variability of physiological age indicators. However, confounding this variability in many age estimation applications is a systematic tendency for age estimates, regardless of physiological indicator employed, to assign ages which are too high for young individuals, and too low for older individuals. This paper shows that at least part of this error is the inevitable consequence of the statistical procedures used to extract an estimate of age from age indicators, and that the magnitude of the error is inversely related to how well an age indicator is correlated with age. The use of classical calibration over inverse calibration is recommended for age estimation. Am J Phys Anthropol 104:259–265, 1997.

Recently there has been a resurgence of interest in one of the oldest anthropological problems, that of adult human age estimation from hard tissue evidence. Important new work concentrates on age estimation for individuals within a relatively large series (Konigsberg and Frankenberg, 1992, 1994). However, there is still a case to be made in favour of the 'forensic' approach in situations where there may be a limited number of individuals available for study, such as the prehistoric tumuli of Western Europe, or whenever the anthropologist needs to know the age of a specific individual. The "forensic" approach should be seen as complementary to the emerging techniques of estimating population age structure.

The "forensic" approach to age estimation has always been seen as problematic because the criteria used to assign age have a high degree of variability within any defined age group; this makes age estimation for an individual inherently imprecise. However, in the literature there are many reports of another problem—an apparently age-dependent systematic error.

Bedford et al. (1993, their Table 2), when testing the Gilbert-McKern age estimation system based on symphaseal changes in the pubic bone, and Wegener and Albrecht (1980, their Figure 3), employing an age estimation technique based upon root dentine translucency, found that estimated ages were too old for young individuals and too young for old individuals. The same systematic discrepancy was also seen in an empirical study of different dental-based age estimation methods (Solheim and Sundnes, 1980).

Saunders et al. (1992, their Table 1) tested the "complex method" for age estimation on documented specimens from Hamilton, Ontario, and Molleson (1993) tested the "complex method" on the skeletal material from

*Correspondence to: D. Lucy, Department of Archaeological Sciences, University of Bradford, Bradford, West Yorkshire, U.K. BD7 1DP. E-mail: d.j.lucy@bradford.ac.uk

Received 20 October 1995; accepted 1 July 1997.

Spitalfields, London. Both these studies noticed the same systematic error seen in the studies cited above.

Masset (1989, p. 81) also noticed this phenomenon and called it "the attraction of the middle." He ascribed it to the skewed distribution of age in the reference sample. Masset tried solving the problem by varying the age structure of the reference sample, but the results were disappointing.

The fact that this systematic error always takes on the same form and can be seen to occur irrespective of the age estimation technique employed suggests the cause lies deeply embedded in the methodology of age estimation. We show that at least part of the problem is the result of using conventional least squares regression to produce age estimates.

LEAST SQUARES REGRESSION AND INVERSE CALIBRATION

Least squares regression analysis arose in the 1880s from Sir Francis Galton's work on heredity (MacKenzie, 1981) and since then has been used for making estimates of one quantity from another in all fields of the physical and social sciences.

Regression analysis is a means of establishing a relationship, expressed as a mathematical equation, between variables which are thought to be related. The simplest form of regression is where there are only two variables and the relationship is assumed to be linear (simple linear regression). A linear relationship between two variables x and y can be expressed by the simple equation:

$$y = a + bx + e \qquad (1)$$

where x is any x-value; y is the corresponding y-value; a is the value taken by y when x equals zero; b is the slope or rate of increase in y with x; and e is the random error in y. It should be noted that the random error is assumed to be entirely in the y direction, which for simple linear regression would be in the age direction.

In most scientific applications the researcher is never given the values of a and b, but instead has to estimate them from data. This is what least squares regression was designed to do. To illustrate this, consider Figure 1, which shows three hypothetical points labelled (x_1, y_1), (x_2, y_2) and (x_3, y_3) where y represents age and x an age indicator. The objective of linear regression is to fit a line $y = a + bx$ in such a way that the sum of the squares of the vertical distances, e_1, e_2, and e_3, is minimised. The squared term is employed because some distances, such as e_1, are negative and some, like e_2 and e_3, are positive, yet we wish to have the total distance between the line and the points minimised irrespective of whether the distance is positive or negative.

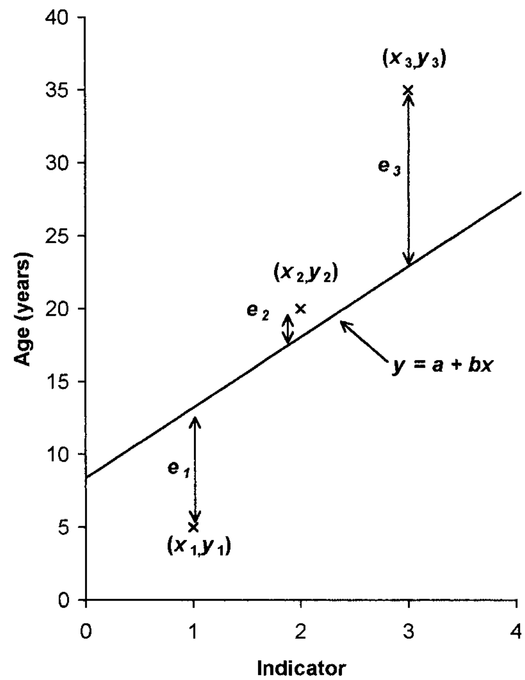

Fig. 1. Part of a regression function.

Here once the regression calculation has been performed it would be considered rational to use them to estimate age for new individuals, a procedure known to statisticians as inverse calibration. For example, in Figure 2 the axes used in Figure 1 have been expanded to show the full dataset used for the regression (it should be noted that these data are only simulated values for illustration purposes). However, were one to do this for an entire sample of known age, then compare the calculated ages to the true ones, a systematic, but biased, relationship would be found. This is shown in Figure 3

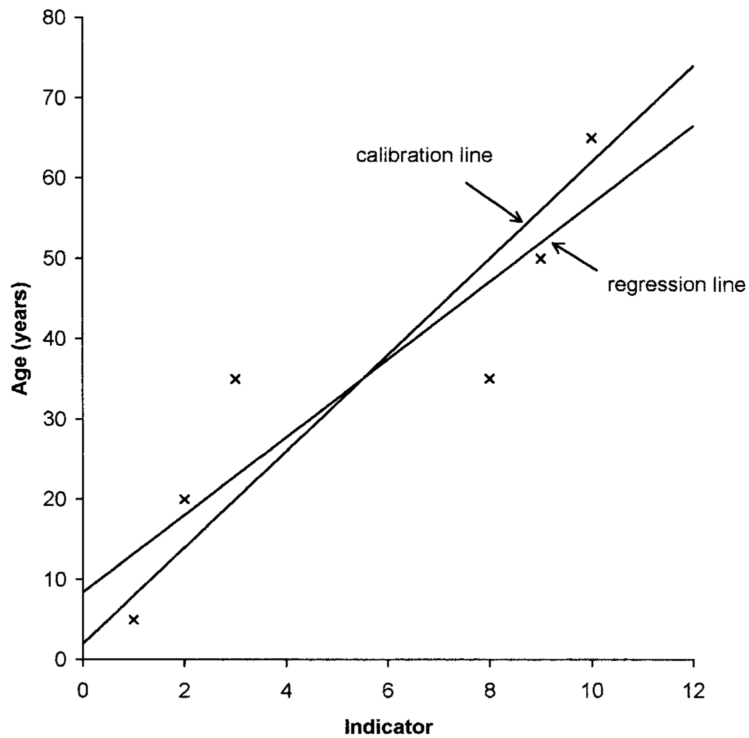

Fig. 2. Regression showing regression and calibration lines.

where the errors (difference between each inverse calibration estimate of age and its corresponding known age) have been plotted against known age. A regression of these errors (the residuals) on known age (the response variable) illustrates this systematic bias. The least squares approach is the same as above, with a change of variable names. The proposed linear relationship is $\epsilon_1 = c + dy$. The mean of the new residuals equals zero, but the slope of this regression line is what is of interest. The slope is equal to $1 - r^2$, where r is the Pearson correlation coefficient between known age and indicator state, from the first regression (see Appendix for derivation). From Figure 3, the graph of the residuals against known age, we can see that the slope is positive; for small values the residual will be negative, and for large values positive. When used for age estimation this guarantees all young individuals will appear older than they really are and all older individuals younger, whatever the dataset.

As the correlation coefficient can only take values between minus one and one this means that the slope of the regression residual can only vary between zero and one. If there were a perfect correlation, with $r^2 = 1$, then a plot of the residuals against the predicted values would reveal a zero slope. Conversely, were there zero correlation, then the slope would equal unity. Put simply, the poorer the correlation the greater the slope. With relatively low correlations between age indicators and age, for most indicators, it is little wonder that anthropologists have been finding their residuals to be correlated with age.

CLASSICAL CALIBRATION

Recently (Konigsberg et al., 1994; Lucy and Pollard, 1995) it has been recommended that a variation of regression, called classical calibration, may be a more suitable statistical technique for making age estimates from adult human age indicators.

Classical calibration is exactly the same as regression analysis except in one important respect. The variable for which estimates are to be made is always x, not y as in regression analysis. In the example, this

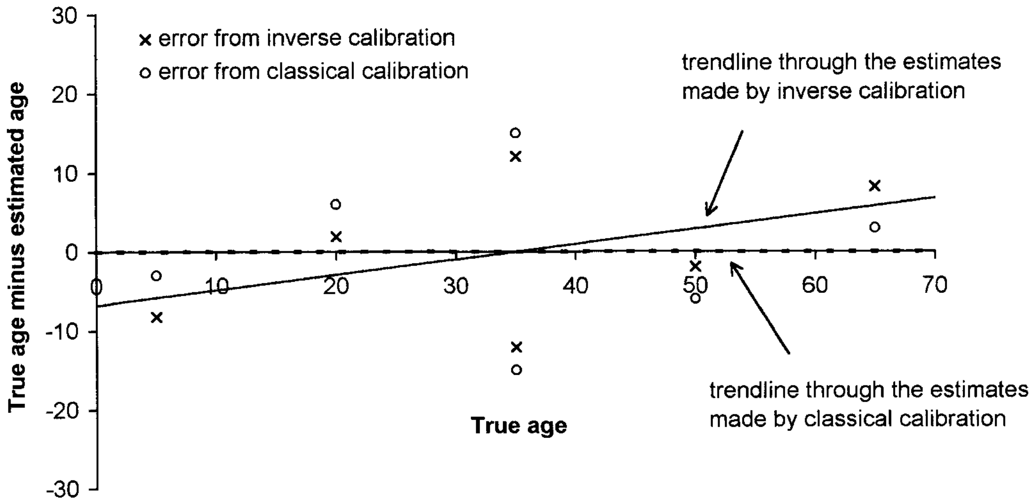

Fig. 3. Plot of inaccuracy in age estimation against age.

would mean that age is x and the indicator value is y. This is shown for the hypothetical example in Figure 2, but the axes have been changed around so the line can be plotted on the same Figure as the regression line. In calibration y is observed and we must estimate x, the reverse of the usual situation. The procedure is widely used for controlled experiments where the x variable is said to be "exactly determined," that is, can be said to be known without error because it is under the control of the experimenter. An example would be the extension in a spring in response to some applied weight. In this case the applied weight is said to be "exactly determined" as it is selected by the experimenter. The response (y) is known with some random error. A regression y on x is performed as usual (Equation 1). This produces an equation for y in terms of x, so to estimate x we must invert this relationship. If the regression equation from the reference sample is $y = a + bx$ then the equation used to estimate x from y is:

$$\hat{x}_i = \frac{(y_i - \hat{a})}{\hat{b}} + e \qquad (2)$$

where y_i is the observation of indicator; $\hat{x}_i$ is the estimate of age for the given indicator; $\hat{a}$ is the estimated intercept from the regression; $\hat{b}$ is the estimated slope from the regression; and e is the random error about $\hat{x}_i$.

Figure 3 shows a regression line through the errors obtained from calibration for the example in Figure 2. The line has been plotted as a dashed line to distinguish it from the x axis as it follows exactly the x axis. This shows that the bias seen in estimates from regression analysis has been completely removed.

The use of classical calibration has some appealing points over regression in practice (Lucy and Pollard, 1995). Least squares regression adjusts the regression equation solely in the y direction making the tacit assumption that all x measurements are made without error, and it is errors in the y-variable which are responsible for any lack of conformity to a straight line relationship. This is the origin of the x-variable being interpreted as the independent variable, and the y-variable as the dependent variable, where the x-variable is somehow seen as controlling the y. Although it is not strictly true that x must always be the controlling variable and y the response, it is true that x must be the variable which is considered to fixed and y to vary about x. In situations where one is trying to derive an expression

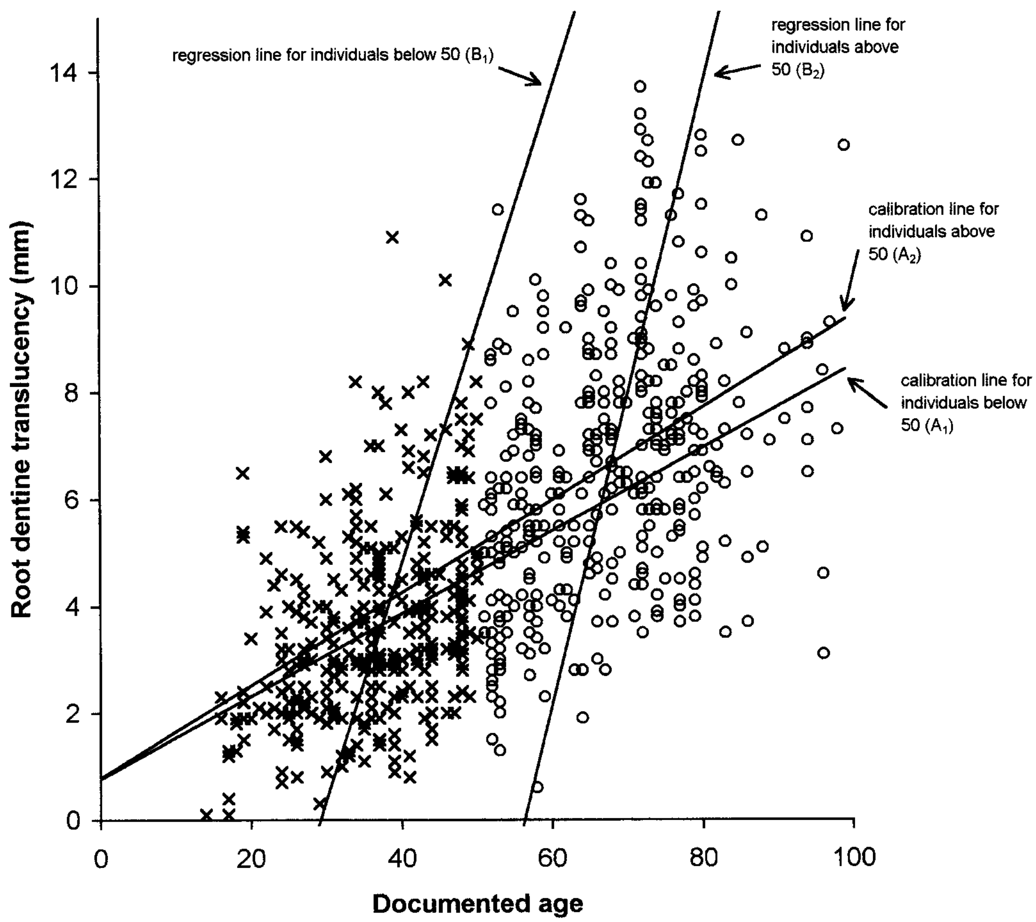

Fig. 4. Age and indicator regression and calibration for a sample partitioned at age = 50 years.

for age from observational data it may be more intuitively appealing to consider age as the variable which is fixed (although there will be some small error due to sampling variation), i.e., independent, and the observation of some age related change as having some random component for any *x*, i.e., dependent. A more biological statement of this interpretation would be "age causes physiological change."

Unfortunately there are a number of drawbacks to calibration. The first is that the uncertainty for any point about the calibration line is rather difficult to calculate (Fisch and Strehlau, 1993), and comprises an expression which describes a confidence interval which is parabolic about the calibration line. This is because the error is in fact a combination of errors from two sources. First, for any indicator score there are a range of possible ages due to biological uncertainty. Second, the estimates of the calibration parameters *a* and *b* also have an uncertainty associated with them which increases with distance from the mean of both age and indicator.

Another drawback is a reduction in the efficiency of estimates (Konigsberg et al., 1994); i.e., the variability will be larger for calibration than for regression, for example in the hypothetical example from Figure 2, the mean inaccuracy for estimates made from regression is 7.4 years, but for calibration it is 8 years.

CLASSICAL CALIBRATION AND A 'FAREWELL TO PALAEODEMOGRAPHY'

In the light of the foregoing discussion it is now possible to re-examine a statistical claim made in a classic paper by Bouquet-Appel and Masset (1982), and which led them to believe that palaeodemography was simply impossible.

Bouquet-Appel and Masset (1982) argued that estimated ages are crucially dependent on the age structure of the reference sample of indicators and age. They suggested that if there were a 'Gaussian partition' of age into two subsets of a reference sample, the result would be two very different regression functions from each subset. They illustrate this with their Figure 2 (i and ii), which depicts regression functions derived from two different subsets for cranial suture data and age. On it are marked functions A_1, A_2, B_1 and B_2. A_1 and A_2 are the calibration lines where the age indicator is *y*, B_1 and B_2 are the regression lines where *y* is age. We present Figure 4 which is a similar plot of indicator and age, although this time using root dentine translucency measurements from previously published data (Solheim, 1989). The sample has been arbitrarily divided at the age of 50 years, regression and calibration lines have been calculated and plotted on the graph for both age subsets, the various lines having been labelled following Bouquet-Appel and Masset's Figure 2.

Indeed functions B_1 and B_2 have very different intercepts with the *y* axis, and as they rightly point out would lead to the effect whereby estimates of age were a reflection of the age structure of the reference sample. Bouquet-Appel and Masset point out that A_1 and A_2 would be suitable for age estimation, but state that it is necessary to use functions which estimate *y* as age on *x* as indicator. We contend that it is really the calibration expressions which should be employed for age estimation (A_1 and A_2), not the regression lines (B_1 and B_2). If this is the case then the model used for age estimation, and the age estimates produced from that model, will not be completely dependent on the age distribution of the reference population in the way suggested by Bouquet-Appel and Masset.

CONCLUSIONS

Many anthropologists when testing "forensic" human age estimation techniques have noticed a systematic, apparently age related, error in the estimates. Part of the reason for this is that regression analysis has been employed to generate the estimates of age, not calibration. Calibration is very similar to regression, but the variable for which estimates are to be made is always treated as *x*. The resultant regression equation then has to be manipulated to produce estimates of *x*, and the estimates will be less efficient than those made by regression. However they will lack the bias in estimation apparent with regression analysis.

ACKNOWLEDGMENTS

The authors thank Lyle Konigsberg and Kathleen O'Connor for their advice in the preparation of this paper, and acknowledge the assistance of an anonymous reviewer for their comments on an earlier draft.

LITERATURE CITED

Bedford ME, Russell KF, Lovejoy CO, Meindl RS, Simpson SW, and Stuart-Macadam PL (1993) Test of the multifactorial aging method using skeletons with known ages-at-death from the Grant collection. Am. J. Phys. Anthropol. *91:*287–297.

Bouquet-Appel JP, and Masset C (1982) Farewell to palaeodemography. J. Hum. Evol. *11:*321–333.

Draper NR, and Smith H (1981) Applied Regression Analysis. 2nd ed. New York: John Wiley.

Fisch RD, and Strehlau GA (1993). A simplified approach to calibration confidence sets. Am. Statistician *47:*168–171.

Konigsberg LW, and Frankenberg SR (1992) Estimation of age structure in anthropological demography. Am. J. Phys. Anthropol. *89:*235–256.

Konigsberg LW, and Frankenberg SR (1994) Paleodemography: "Not quite dead." Evol. Anthropol. *3:*92–105.

Konigsberg LW, Frankenberg SR, and Walker RB (1994) Regress what on what?: Paleodemographic age estimation as a calibration problem. In RR Paine (ed.): Integrating Archaeological Demography: Multidisciplinary Approaches to Prehistoric Population. Carbondale: Southern Illinois University, pp. 64–88.

Lucy D, and Pollard AM (1995) Further comments on the estimation of error associated with the Gustafson dental age estimation method. J. Forensic Sci. *40:*222–227.

MacKenzie DA (1981) Statistics in Britain 1865–1930: The social construction of scientific knowledge. Edinburgh: Edinburgh University Press.

Masset C (1989) Age estimation on the basis of crainial sutures. In MY Iscan (ed.): Age Markers in the Human

Skeleton. Springfield, Illinois: Charles C. Thomas, pp. 71–103.
Molleson T (1993) Growing old. In T Molleson, M Cox, AH Waldron, and DK Whittaker (eds.): The Spittlefields Report Volume 2, The Anthropology: The Middling Sort. Council for British Archaeology: Research Report Number 86.
Saunders SR, Fitzgerald C, Rogers T, Dudar C, and McKillop H (1992) A test of several methods of skeletal age estimation using a documented archaeological sample. Can. Soc. Forensic Sci. *25:*97–118.
Solheim T (1989) Dental root transparency as an indication of age. J. Dent. Res. *97:*189–197.
Solheim T, and Sundnes PK (1980) Dental age estimation of Norweigian adults: A comparison of different methods. Forensic Sci. Int. *16:*7–17.
Wegener R, and Albrecht H (1980) Estimation of age from root dentine transparency. Z. Rechtsmed. *87:* 29–34.

APPENDIX

In any least squares linear regression a regression function is derived which describes the least squares relationship between y and x.

$$y_i = \hat{a} + \hat{b}x_i = \epsilon_i, \quad i = 1, \ldots, n \tag{3}$$

with the least squares estimated values for $\hat{a}$ and $\hat{b}$:

$$\hat{b} = \frac{S_{xy}}{S_{xx}} \tag{4}$$

and

$$\hat{a} = \bar{y} - \hat{b}\bar{x}. \tag{5}$$

where

$$S_{xy} = \frac{1}{n}\sum (x_i - \bar{x})(y_i - \bar{y})$$

and

$$S_{xx} = \frac{1}{n}\sum (x_i - \bar{x})^2.$$

If we now consider the residuals which are obtained by subtracting the estimated y-value from the observed y-value:

$$\hat{\epsilon}_i = y_i - \hat{y}_i,$$

and from Equation 3 and Equation 5:

$$\hat{\epsilon}_i = y_i - (\bar{y} - \hat{b}\bar{x} + \hat{b}x_i) = y_i - \hat{y} - \hat{b}(x_i - \bar{x}),$$

and from Equation 4:

$$\hat{\epsilon}_i = y_i - \bar{y} - \frac{S_{xy}}{S_{xx}}(x_i - \bar{x}).$$

To examine any relationship between these residuals and the response variable, y, we shall regress the residuals on the response variable. The least squares approach is the same as above, with a change of variable names. Let the proposed linear relationship be $\hat{\epsilon}_i = c + dy_i$; then, using least squares:

$$\hat{c} = \bar{\hat{\epsilon}} - \hat{d}\bar{y},$$

and

$$\hat{d} = \frac{\Sigma(y_i - \bar{y})(\bar{\hat{\epsilon}}_i - \bar{\epsilon})}{\Sigma(y_i - \bar{y})^2}.$$

It is easily shown that the mean of the new residuals equals zero. However, what we are really interested in is the slope of the regression line ($\hat{d}$) and, more specifically, whether it is non-zero.

$$\begin{aligned}\hat{d} &= \sum \frac{(y_i - \bar{y})}{\Sigma(y_i - \bar{y})^2}\left\{y_i - \bar{y} - \frac{S_{xy}}{S_{xx}}(x_i - \bar{x})\right\} \\ &= \frac{\Sigma(y_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} - \frac{S_{xy}}{S_{xx}}\frac{\Sigma(y_i - \bar{y})(x_i - \bar{x})}{\Sigma(y_i - \bar{y})^2} \\ &= 1 - \frac{S_{xy}}{S_{xx}}\frac{S_{xy}}{S_{yy}} \quad = 1 - r_{xy}^2\end{aligned}$$

where r_{xy} is the Pearson product-moment correlation coefficient between x and y.

This result, but not the derivation, also appears in the second edition of Draper and Smith (1981).