THE SKEWNESS OF THE RESIDUALS IN LINEAR REGRESSION THEORY

By P. S. Dwyer

University of Michigan, Ann Arbor, Mich.

In obtaining the regression of y on x it is customary to show the relation between the actual and the estimated y by computing the standard deviation of the residuals with the use of the formula $\sigma_\epsilon = \sigma_y \sqrt{1 - r^2}$. If the errors are distributed normally one may estimate the number of values coming within one standard deviation, within two standard deviations, etc., of the regression line. However these errors are not always distributed normally, and in such a case it seems wiser to compute the skewness of the residuals and to use a Pearson Type III curve in making the interpretation. The present paper outlines a technique for the calculation of $\alpha_{3:\epsilon}$ which is feasible from a practical standpoint. It is based (*a*) on a cumulative totals method of obtaining the correlation coefficient which, at the same time, makes possible the determination of the third order moments needed to evaluate the skewness and (*b*) on an efficient ritual for computing the coefficient of skewness from the moments.

The determination of the normality or non-normality of the residuals is not always immediately evident. If the scatter diagram or correlation chart is presented, one can make an estimate of the extent of normality but if not, and the most modern and efficient computational methods do not utilize the correlation chart, there is no way by which the presence or absence of normality can be detected. Some research workers are opposed to the use of the more efficient methods (particularly the use of the Hollerith tabulators) because the correlation chart is not presented. Though within limits it is possible to use the tabulator to present the correlation chart simultaneously with the values needed to compute the correlation coefficient [1], it is here suggested that the computation of the skewness of residuals, which can now be accomplished quite easily from the tabulator runs, may be substituted for the examination of the correlation chart.

The classical least squares theory makes use of

$$\epsilon = y - b_0 - b_1 x \tag{1}$$

where b_0 and b_1 are the solutions of the normal equations. We note that the first normal equation is $\Sigma\epsilon = 0$ so that $M_\epsilon = 0$ and the residual is a deviation. It follows that the skewness of residuals is

$$\alpha_{3:\epsilon} = \frac{\Sigma(y - b_0 - b_1 x)^3}{N\sigma_\epsilon^3}. \tag{2}$$

We wish to compute $\alpha_{3:\epsilon}$ without computing the individual residuals. The denominator causes us little concern but it seems discouraging to evaluate such an expression as

$$\Sigma y^3 - Nb_0^3 - b_1^3\Sigma x^3 - 3b_0\Sigma y^2 - 3b_1\Sigma xy^2 + 3b_0^2\Sigma y - 3b_0^2b_1\Sigma x + 3b_1^2\Sigma x^2y - 3b_1^2b_0\Sigma x^2 + 6b_0b_1\Sigma xy$$

even though the values of b_0, b_1, N, Σx, Σy, Σx^2, Σxy, Σy^2, Σx^3, Σx^2y, Σxy^2, Σy^3 are available.

A first simplification is made by summing (1) and dividing by N. We then have

$$M_\epsilon = M_y - b_0 - b_1M_x \tag{3}$$

and by subtracting (3) from (1) and denoting deviations by barred letters, we have

$$\epsilon = \bar{y} - b_1\bar{x} \tag{4}$$

so that the skewness of errors is

$$\alpha_{3:\epsilon} = \frac{\Sigma\bar{y}^3 - 3b_1\Sigma\bar{x}\bar{y}^2 + 3b_1^2\Sigma\bar{x}^2\bar{y} - b_1^3\Sigma\bar{x}^3}{N\sigma_\epsilon^3}. \tag{5}$$

This formula can also be expressed as

$$\alpha_{3:\epsilon} = \frac{\bar{\mu}_{03} - 3b_1\bar{\mu}_{12} + 3b_1^2\bar{\mu}_{21} - b_1^3\bar{\mu}_{30}}{[\bar{\mu}_{02} - b_1\bar{\mu}_{11}]^{3/2}}. \tag{6}$$

A similar formula for the skewness of the residuals of x on y is

$$\alpha_{3:\epsilon'} = \frac{\bar{\mu}_{30} - 3b_1'\bar{\mu}_{21} + 3b_1'^2\bar{\mu}_{12} - b_1'^3\bar{\mu}_{03}}{[\bar{\mu}_{20} - b_1'\bar{\mu}_{11}]^{3/2}}. \tag{7}$$

For theoretical purposes formula (6) may be put in standard units with $b_1 = r\frac{\sigma_y}{\sigma_x}$, $b_1' = r\frac{\sigma_x}{\sigma_y}$, $\bar{\mu}_{30} = \alpha_{30}\sigma_x^3$, $\bar{\mu}_{21} = \alpha_{21}\sigma_x^2\sigma_y$, etc. with the resulting

$$\alpha_{3:\epsilon} = \frac{\alpha_{03} - 3r\alpha_{12} + 3r^2\alpha_{21} - r^3\alpha_{30}}{(1 - r^2)^{3/2}}. \tag{8}$$

As $r \to 0$, $\alpha_{3:\epsilon} \to \alpha_{3:y}$ just as $\sigma_\epsilon \to \sigma_y$ as $r \to 0$.

Formulas (6) and (7) are of some theoretical importance in that they show how the skewness of the residuals is connected with the skewness of the marginal distribution. Thus

as $\bar{\mu}_{11} \to 0$, b_1 and $b_1' \to 0$ and $\alpha_{3:\epsilon} \to \alpha_{3:y}$, $\alpha_{3:\epsilon'} \to \alpha_{3:x}$;

as $b_1 \to \infty$, $\alpha_{3:\epsilon} \to -\alpha_{3:x}$ and as $b_1' \to \infty$, $\alpha_{3:\epsilon'} \to -\alpha_{3:y}$;

as $b_1 \to 1$, $\alpha_{3:\epsilon} \to \alpha_{3:y-x}$. Similarly as $b_1' \to 1$, $\alpha_{3:\epsilon'} \to \alpha_{3:x-y}$.

It is hence possible in some cases to get a good approximation to the skewness of the residuals if the regression coefficients and the skewness of the marginal distribution are known.

TABLE I
Correlation from first order cumulations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
$Y \backslash X$			4.00	3.99 3.50-	3.49 3.00-	2.99 2.50-	2.49 2.00-	1.99 1.50-	1.49 1.00-	.99 .50-	.49 .00-		
	$y \backslash x$		8	7	6	5	4	3	2	1	0		
		$f_y \backslash f_x$	13	50	107	220	341	179	121	60	35	Cx_y	Cy_y
4.00	6	18	5	2	5	5	1					113	108
3.99 3.50-	5	106	2	19	29	27	20	7		1	1	673	638
3.49 3.00-	4	178	3	12	35	53	44	18	6	5	2	1503	1350
2.99 2.50-	3	270	3	10	20	55	103	33	27	11	8	2568	2160
2.49 2.00-	2	330		6	11	54	114	67	46	19	13	3714	2820
1.99 1.50-	1	173		1	5	19	45	44	34	18	7	4244	2993
1.49 1.00-	0	51			2	7	14	10	8	6	4	4399	2993
		Cy_x	61	259	661	1330	2194	2578	2809	2923	2993	12815	10069
		Cx_x	104	454	1096	2196	3560	4097	4339	4399	4399	20245	

For actual computation, we use (6) and (7). It has been indicated previously how the values Σx, Σy, Σx^2, Σxy, Σy^2, Σx^3 and Σy^3 could be obtained with the use of cumulations. An illustration used previously [2] is presented in Table I. The information was obtained from the Office of Educational Investigations of the University of Michigan and gives the University first semester average (X) and the high school average (Y) for 1,126 students entering the College of Literature, Science, and the Arts in 1928.

The new origin of each variable is taken at the class mark of the lowest class rather than at the class mark of a middle class as is conventional. In this way all negative terms are avoided in the computation of the moments. The x's are arranged in descending order from left to right and the y's in descending order from top to bottom. The notation x_y is used to indicate the sum of all the x's having the same value of y. Thus the first entry in column 13 is $5\cdot 8 + 2\cdot 7 + 5\cdot 6 + 5\cdot 5 + 1\cdot 4 = 113$. The column Cx_y is obtained by cumulating the values of x_y. Similarly y_y is the sum of all the y's having the same value and the first entry in column 14 is $18(6) = 108$. The entries Cy_y, Cy_x, and Cx_x are obtained similarly.

The entries Σx, Σy, Σx^2, Σxy, Σy^2 are found in the lower right hand box in this position:

	Σx	Σy
Σy	Σxy	Σy^2
Σx	Σx^2	

The values of Σx and Σy are obtained from the final cumulations while the value of Σxy is obtained by adding the entries in the column above, or, as a check, the entries in the row to the left. The value of Σy^2 is obtained by adding the entries in the row at the left of the box while the value Σx^2 is obtained by adding the entries above the box.

The values of the third order sums are obtained by multiplying the entries above the box and to the left of the box successively by 1, 3, 5, 7, 9, etc. Thus,

$$\begin{aligned}
\Sigma x^3 &= 4399 + 3(4339) + 5(4097) + \text{etc.} = 102{,}103, \\
\Sigma x^2 y &= 2923 + 3(2809) + 5(2578) + \text{etc.} = 63{,}121, \\
\Sigma xy^2 &= 4244 + 3(3714) + 5(2568) + \text{etc.} = 46{,}047, \\
\Sigma y^3 &= 2993 + 3(2820) + 5(2160) + \text{etc.} = 38{,}633.
\end{aligned} \tag{9}$$

In making the reductions we use $ab - cd$ operations as much as possible. We first compute

$$\begin{aligned}
A_{x,y} &= N\Sigma xy - (\Sigma x)(\Sigma y), \\
A_{x,x} &= N\Sigma x^2 - (\Sigma x)^2, \\
A_{x^2,y} &= N\Sigma x^2 y - (\Sigma x^2)(\Sigma y).
\end{aligned} \tag{10}$$

We note too that

$$\begin{aligned}
\bar{\mu}_{30} &= [NA_{x^2,x} - (2\Sigma x)(A_{x,x})]/N^3; \qquad \bar{\mu}_{21} = [NA_{x^2,y} - (2\Sigma x)(A_{x,y})]/N^3 \\
\bar{\mu}_{12} &= .[NA_{x,y^2} - (2\Sigma y)(A_{x,y})]/N^3; \qquad \bar{\mu}_{03} = [NA_{y^2,y} - (2\Sigma y)(A_{y,y})]/N^3
\end{aligned} \tag{11}$$

and finally we get $\alpha_{3:\epsilon}$ or $\alpha_{3:\epsilon'}$ by (6) or (7).

The general solution is outlined on the left of Table II. We record in Fig. A the values given by (9) and in the Fig. B the values resulting from the application of (10). The values $2\Sigma y$ and $2\Sigma x$ are inserted in Fig. B to facilitate the calculation of Fig. C which gives the values of (11). The technique is very easily carried out once it is understood. It can be performed with hand calculators but it is ideally adapted to the use of the latest Marchant, Fridén, and Monroe models equipped with automatic positive and negative multiplication, so that *ab–cd* operations can be performed with a minimum of effort and a maximum of accuracy. Actually the value of "*a*," which is the total frequency, is the same for many of these operations so that there is further saving if a machine is used which permits the locking in of a constant in such a way that it can be used, without continued key punching, in later ab-cd operations.

TABLE II

Abbreviated techniques for computing third order central moments, etc.

Fig. A.

N	Σx	Σx^2	Σx^3	1126	4399	20245	102103
Σy	Σxy	$\Sigma x^2 y$		2993	12815	63121	
Σy^2	Σxy^2			10069	46047		
Σy^3				38633			

Fig. B.

N	$2\Sigma x$	$A_{x,x}$	$A_{x^2,x}$	1126	8798	3444669	25910223
$2\Sigma y$	$A_{x,y}$	$A_{x^2,y}$		5986	1263483	10480961	
$A_{y,y}$	$A_{y^2,x}$			2379645	7555391		
$A_{y^2,y}$				13364241			

Fig. C.

N		$A_{x,x}$	$N^3\bar{\mu}_{30}$	1126		3444669	−1131286764
	$A_{x,y}$	$N^3\bar{\mu}_{12}$			1263483	685438652	
$A_{y,y}$	$N^3\bar{\mu}_{21}$			2379645	944161028		
$N^3\bar{\mu}_{03}$				803580396			

Fig. D.

N	(b_1)	$\bar{\mu}_{20}$	$\bar{\mu}_{03}$	1126	(.367)	2.717	−.7925
(b_1')	$\bar{\mu}_{11}$	μ_{21}	$(-b_1^3)$, $(-3b_1')$	(.531)	.997	.4801	(−1.593)
$\bar{\mu}_{02}$	$\bar{\mu}_{12}$	$(3b_1^2)$, $(3b_1'^2)$		1.877	.6614	(.846)	
$\bar{\mu}_{03}$	$(-3b_1)$, $(-b_1'^3)$			.5629	(−.150)		

The values in Fig. D are obtained by dividing the values $A_{y,y}$, $A_{x,y}$, and $A_{x,x}$ in Fig. C by N^2 and the values in the diagonal below, $NA_{y^2,y} - (2\Sigma y)A_{y,y}$, etc., by N^3. The values $b_1 = \frac{\bar{\mu}_{11}}{\bar{\mu}_{20}}$ and $b_1' = \frac{\bar{\mu}_{11}}{\bar{\mu}_{02}}$ can be inserted in Fig. D adjacent to the N. The value of the correlation coefficient is $r = \sqrt{b_1 b_1'} = \frac{\bar{\mu}_{11}}{\sqrt{\bar{\mu}_{20}\bar{\mu}_{02}}}$.

We have too, $\sigma_\epsilon = \sqrt{\bar{\mu}_{02} - b_1\bar{\mu}_{11}}$ and $\sigma_{\epsilon'} = \sqrt{\bar{\mu}_{20} - b_1'\bar{\mu}_{11}}$ so that the standard deviation of residuals is readily computed from the entries of Fig. D. The numerator of (6) is readily obtained after entering $-3b_1$, $3b_1^2$, $(-b_1^3)$ in the diagonal under the diagonal containing the third moments and multiplying by columns. The numerator of (7) is obtained by entering $-b_1'^3$, $3b_1'^2$, $-3b_1'$, in the same diagonal and multiplying by rows. The theory is applied to the results of Table I and the details are presented at the right of Table II. It is to be noted that all values indicated here are the coded values x, y and not the original values, X, Y. However, the correlation coefficient and the skewness of errors are independent of any such change in unit, grouping errors being neglected.

From Fig. D we see that $b_1 = .997/2.717 = .367$, that $b_1' = .997/1.877 = .531$ and that $r = \sqrt{(.367)(.531)} = .441$. In this case we wish to estimate college record, x, from high school record, y, so we use $b_1' = .531$ and compute $-3b_1' = -1.593$, $3b_1'^2 = .846$, $-b_1'^3 = -.150$. It follows that

$$\alpha_{3:\epsilon'} = \frac{-.7925 + (.4801)(-1.593) + (.6614)(.846) + (.5629)(-.150)}{[2.717 - .531(.997)]^{3/2}} = -.334.$$

It thus appears that a better picture of the variation of the residuals in this case is obtained with the use of a Pearson Type III with α_3 approximately $-\frac{1}{3}$ than is obtained with the use of a normal curve. It is not necessary, of course, to form Fig. D as the results can all be obtained from Fig. C. Thus if we multiply the numerator and denominator of (6) by N^3, we get entries, with the exception of the b's, which are in Fig. C. Now in this case $b_1 = \dfrac{A_{x,y}}{A_{x,x}}$ and $b_1' = \dfrac{A_{x,y}}{A_{y,y}}$ so that these values can be inserted in the upper left as before. Also the powers of b_1' can be inserted in the lower right as in Fig. D. We have then

$$\alpha_{3:\epsilon'} = \frac{\begin{array}{c}-1131{,}286{,}764 + (685{,}438{,}652)(-1.593) + (944{,}161{,}028)(.846)\\ + (803{,}580{,}396)(-.150)\end{array}}{[3444669 - (1263483)(.531)]^{3/2}}.$$

We know however, since the grades were coded, that it is not sensible to carry results to more than three places, (and, indeed, a three place determination of the skewness is very satisfactory for interpretive purposes even though more places might be obtained) so we cut down the number of places. The division of numerator and denominator by 10^6, and the dropping of the decimals results in

$$\alpha_{3:\epsilon'} = \frac{-1131 + 685(-1.593) + 944(.846) + 804(-.150)}{[344 - 126(.531)]^{3/2}} = -.335.$$

It is possible of course to duplicate the theory indicated in Table II with the use of moments rather than the A's. In this case Fig. A consists of 1, $\Sigma x/N$,

$\Sigma x^2/N$, etc. We have such formulas as $a_{x,y} = \frac{\Sigma xy}{N} - \frac{(\Sigma x)}{N}\frac{(\Sigma y)}{N} = \mu_{11} - \mu_{10}\mu_{01}$, where $a_{xy} = \frac{A_{x,y}}{N_2}$, $a_{x^2,y} = \frac{A_{x^2,y}}{N_3}$, etc.

It would be possible to compute the $\alpha_{4:\epsilon}$ in a somewhat similar fashion though it would take somewhat longer. In the first place we would have to compute Σx^2y^2 from the correlation table. This could be done by forming the cumulation $C(y_x^2)$ and multiplying by 1, 3, 5, 7, 9, etc. When this is done, however, it does not appear that the calculation of the central moments of the fourth order can be reduced to as simple a ritual as the calculation of the third order moments.

The question should be raised as to the calculation of the skewness when there are two or more independent variables. This can be done, of course, but the calculations are lengthy. The point of the present paper is to provide an easy and simple technique for computing the skewness of residuals in the case of two variable linear regression.

REFERENCES

[1] Paul S. Dwyer and Alan D. Meacham, "The preparation of correlation tables on a tabulator equipped with digit selection," *Jour. Am. Stat. Ass.*, Vol. 32 (1937), pp. 654–62. See particularly page 657.

[2] Paul S. Dwyer, "The computation of moments with the use of cumulative totals," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 288–304. See particularly pages 299-303.