ON A NEW METHOD OF COMPUTING NON-LINEAR REGRESSION CURVES*

By

WALTER ANDERSSON,
fil. dr, Stockholm.

In a memoir published in this journal in February 1930[1] Professor S. D. Wicksell pointed out that the well-known Pearson method[2] of computing skew regression curves by adopting the principle of least squares can be simplified, and in some direction generalized, by inserting some assumption concerning the distribution function of the population studied. After some remarks on the subject as advanced in the said memoir the problem was presented to me by Professor Wicksell. The results obtained by me as regards this problem were published as a part of my doctor thesis.[3] In the course of the official ventilation of my thesis Professor Wicksell made some interesting remarks concerning the relations between my solution and the general Pearson solution. His suggestion has led me to take up this special problem, which will be considered in the following lines.

I. We consider a bi-variate distribution and denote the variables x and y. The distribution function—for the sake of sim-

* From the Statistical Institution of the University of Lund, Sweden.

[1] *S. D. Wicksell,* Remarks on Regression.

[2] *Karl Pearson,* On the General Theory of Skew Correlation and Non-Linear Regression; Mathematical Contributions to the Theory of Evolution / Drap. Comp. Res. Mem., Biom. Ser. II, 1905.

[3] *Walter Andersson,* Researches into the Theory of Regression, chapters IV-VI, / Kungl. Fysiografiska Sällskapets Handlingar, N. F. Bd. 43, Nr. I; also as Meddelande från Lunds Observatorium, Ser. II, Nr. 64 /

plicity being supposed discontinuous—may be

$$(1) \qquad z = F(x, y),$$

so that

$$(2) \qquad \sum_x \sum_y F(x, y) = 1.$$

Let $g(x)$ be the regression function of y on x. Thus

$$(3) \qquad \bar{y}_x = g(x),$$

where $\bar{y}_x$ denotes the mean value of the dependent variate y for a fixed value of the independent variate x. Consequently we have

$$(4) \qquad g(x) = \frac{\sum_y y \cdot F(x,y)}{\sum_y F(x,y)}.$$

We further observe that the marginal distribution of x is

$$(5) \qquad f(x) = \sum_y F(x,y).$$

Expanding the regression function in the series of Tchebycheff we put

$$(6) \qquad g(x) = \alpha_0 \cdot \psi_0(x) + \alpha_1 \cdot \psi_1(x) + \alpha_2 \cdot \psi_2(x) + \cdots,$$

where $\psi_i(x)$ are polynomials of the i^{th} orders, fulfilling the following condition of orthogonality

$$(7) \qquad \sum_x f(x) \cdot \psi_i(x) \cdot \psi_j(x) = 0, \qquad \text{for } i \neq j,$$

and

$$(8) \qquad \sum_x f(x) \cdot \left[g(x) - \alpha_0 \cdot \psi_0(x) - \alpha_1 \cdot \psi_1(x) - \cdots - \alpha_h \cdot \psi_h(x)\right]^2 = \text{Min}.$$

From (7) and (8) it may be shown that the expansion by Tchebycheff carried to some order gives the same approximate expression for the regression as obtained by fitting a parabola of the same order to the mean values of y for every value of x,

[4] *Tchebycheff*, Collected Works, Vol. I, pp. 203-230.

each observation being allotted a weight proportional to the number of individuals possessing the value of x in question. Thus, by using the series of Tchebycheff in treating the regression problem we have as a matter of fact applied the same method of describing the regression as applied by Yule[5] and Pearson.[6]

We observe that using the series of Tchebycheff we gain the advantage of being able to perform the graduation successively for the higher orders. With respect to this circumstance I have used the notation *successive regression coefficients* for the coefficients α_i of (6).

Working out the solution for these coefficients we obtain from (7) and (8),

$$\alpha_i = \frac{\sum_x f(x) \cdot \psi_i(x) \cdot g(x)}{\sum_x f(x) \cdot [\psi_i(x)]^2}, \tag{9}$$

the polynomials $\psi_i(x)$ being determined from (7).

The successive regression coefficients—except α_0 —have been shown / see W. Andersson, Op. cit., pp. 14-15 / to be independent of the zero-values of the variables, and in some cases they are found to stand in simple relations to the well-known semi-invariants of Thiele.[7] Especially when the distribution is assumed to be generated according to the *hypothesis of elementary errors* the semi-invariants of Thiele and the successive regression coefficients are closely related. In this respect the denomination *semi-invariant regression coefficients* may be suggested for the coefficients α_i. The values of these coefficients ought to be derived in all more exhaustive studies of curved regression lines.

---

[5] *G. U. Yule,* On the Significance of Bravais' Formulae for Regression, &c., in case of Skew Correlation / Proc. Roy. Soc., Vol. 60, pp. 477-489, 1897 /.

[6] *Pearson,* Op. cit.

[7] *T. N. Thiele,* Theory of Observations, London 1903, p. 24. / See also Annals of Mathematical Statistics, Vol. II, pp. 165-307, where this work of Thiele is reprinted /.

We introduce the *moments*, ν'_{ij}, of the distribution. Taking these about any point we have

$$(10) \qquad \nu'_{ij} = \sum_x \sum_y x^i \cdot y^j \cdot F(x,y).$$

If we observe that

$$(11) \qquad \sum_x f(x) \cdot x^h \cdot g(x) = \sum_x \sum_y x^h \cdot y \cdot F(x \cdot y),$$

it is immediately seen from (9) that the coefficients α_i can be expressed as linear functions of the "mixed" moments ν'_{01}, ν'_{11}, ν'_{21} up to ν'_{h}, all other quantities being dependent on the marginal moments of x alone.

This solution may shortly be summed up. For a fuller discussion I refer to the cited memoir by the writer.

We write

$$(12) \qquad \psi_i(x) = x^i + e_{i,i-1} x^{i-1} + e_{i,i-2} x^{i-2} + \cdots + e_{i,1} x + e_{i,0}$$

Let $\Delta^{(i)}$ be the following determinant of the marginal moments of x,

$$(13) \qquad \Delta^{(i)} = \begin{vmatrix} 1 & \nu'_{10} & \nu'_{20} & \nu'_{h0} \\ \nu'_{10} & \nu'_{20} & \nu'_{30} & \nu'_{h+1,0} \\ & & & \\ \nu'_{h0} & \nu'_{h+1,0} & \nu_{h+2,0} & \nu'_{2h,0} \end{vmatrix}$$

and $\Delta^{(i)}_{hL}$ be its sub-determinant obtained by cutting out the $(h+1)\,th$ row and the $(L+1)\,th$ column and multiplying by $(-1)^{h+1}$. Then we have

$$(14) \qquad e_{ij} = \frac{\Delta^{(i)}_{ij}}{\Delta^{(i)}_{ii}},$$

and

$$(15) \qquad \alpha_i = \frac{\Delta^{(i-1)}}{\Delta^{(i)}} \left[\nu'_{i1} + e_{i,i-1} \nu'_{i-1,1} + \cdots + e_{i1} \nu'_{11} + e_{i0} \nu'_{01} \right],$$

or, using the "standardized" variables

(16) $$\xi = \frac{x - m_1}{\sigma_1}\ , \quad \eta = \frac{y - m_2}{\sigma_2}\ ,$$

(m = mean, σ = dispersion)

and introducing the coefficients

(17) $$\rho_{i0} = \varepsilon_{i1} - r\ \varepsilon_{i+1,0}$$

where ε_{ij} stands for the "standardized" moments and r is the usual Galton coefficient of correlation, we have / W. Andersson, Op. cit., p. 16 /

(18) $$\alpha_i = \frac{\Delta^{(i-1)}}{\Delta^{(i)}}\left[\rho_{i0} + e_{i,i-1}\,\rho_{i-1,0} + \cdots + e_{i2}\,\rho_{30}\right].$$

The relations between the successive or the semi-invariant regression coefficients α_i and the coefficients of the graduation parabolas as written in their usual forms are easily obtained. Taking the parabola of the p^{th} order

(19) $$\bar{y}_x = a_0^{(p)} + a_1^{(p)} x + a_2^{(p)} x^2 + \cdots + a_p^{(p)} x^p ,$$

we have, indeed, / Op. cit., p. 17 /

(20) $$\begin{aligned} a_0^{(p)} &= \alpha_0 + e_{10}\,\alpha_1 + e_{20}\,\alpha_2 + \cdots + e_{p0}\,\alpha_p \\ a_1^{(p)} &= \alpha_1 + e_{21}\,\alpha_2 + \cdots + e_{p1}\,\alpha_p \\ a_2^{(p)} &= \alpha_2 + \cdots + e_{p2}\,\alpha_p \\ &\cdots\cdots\cdots\cdots \\ a_p^{(p)} &= \alpha_p \end{aligned}$$

The coefficients e_{ij} are the same as those defined by (14).

2. Starting with the general solution just indicated we may

proceed further into the matter. It will be seen that some new problems are met with in applying the general method to actual statistics.

Taking account of the fact that the solution only involves the moments of the distribution, we can free ourselves from any assumptions as regards the distribution function itself. The required moment values may then be directly computed from the observed frequencies. This way of solving the problem leads to the method advanced by Pearson in his treatises on this subject. The solution evidently gives a least squares graduation to the observed array means when the weights of each mean value are proportional to the observed frequencies in the corresponding arrays.

This method may be the most straight-forward one, but it is, however, by no means the simplest, nor the most efficient one. Considering the fact that the term of the i^{th} order of the parabola contains moments up to the $2i^{th}$ order, we immediately conclude that the arithmetical work would rise to a considerable amount, and, with growing moment order be more and more in vain, as a consequence of the rapidly increasing sampling errors of the computed moment values. Some other ways to treat the problem must be sought for in order to eliminate these difficulties.

A first outline of a new method was suggested by S. D. Wicksell in the year 1930 / Wicksell, Op. cit. /. Starting with the general solution Wicksell pointed out that some well-known rules of Thiele as regards the determination of moments of high orders were directly applicable in the computation of high order regression parabolas. The rules of Thiele referred to may be formulated in the following way / Thiele, Op. cit., p. 24 /:

To obtain the first semi-invariants, or moments, rely entirely on computations. To obtain the intermediate semi-invariants rely partly on computations, partly on theoretical considerations. But to obtain the higher semi-invariants rely entirely on theoretical considerations.

Professor Wicksell's suggestion was that instead of the higher marginal moments, involved in the least squares expressions for the regression coefficients, should be inserted the moments of a suitably chosen frequency function (with a limited number of parameters), fitted to the marginal distribution of the independent variate.

The method indicated was then more thoroughly studied by the writer of these lines / Op. cit. /. The solution obtained along these lines was in detail worked out, and it was also tested as regards its practical usefullness in dealing with actual statistics. Especially by use of the Pearson types of frequency functions very simple expressions, successfully applicable within a large domain of actual statistics, were deduced.

An important advantage of this method / as well as of the Pearson method / of computing high order regression parabolas may be noticed. As the regression coefficients have been expressed as functions of the moments only—in the method elaborated by the author only of those of low orders—the influence of "grouping" may be accounted for by correcting the computed moment values in this respect. For this purpose suitable correction formulas are available, as for instance the well-known ones given by Sheppard. Experience has convinced me that at the ends of the regression curves, at least, the effect of grouping can displace the computed curve in a considerable manner, so that in many cases some attention must be paid to these circumstances.

It is, however, to be remembered that the solution obtained by applying Wicksell's proposition does not give a strict least squares graduation to the observed array means, as a consequence of the fact that the theoretical values of the high order moments always in some degree differ from the directly computed ones. From this it is evident that some care must be taken in choosing the hypothesis as regards the marginal distribution of the independent variate. It may be remarked, however, that these circumstances cause very little practical difficulty on account of the much refined theory of uni-variate distributions.

The discrepancy between a least squares solution and the solution as obtained by applying the method as advanced by the author may, as pointed out to me by Professor Wicksell in the course of the official ventilation of my thesis, be removed by an adjustment by which the latter solution is turned into a strict least squares solution. This problem will be considered in the following paragraph, and at the same time we shall get an opportunity to study the hypothetical assumptions applied before from a somewhat different point of view.

3. We consider the expression (8). Before we have from this condition worked out the general least squares solution in assuming $f(x)$ to be the true marginal distribution (5) and $g(x)$ to be the true regression function and *then* 1 / the directly computed moment values were inserted in the general solution / Pearson's method /, or 2 / the moment values required were determined in accordance with the rules indicated in § 2 / method elaborated by the writer /. Now we shall directly imply in (8) our working hypothesis concerning the marginal distribution of x. Let the hypothetical x-marginal distribution function be $\omega(x)$. The solution is then to be deduced from the following condition

$$(21) \qquad \sum_x \omega(x) \cdot \left[g(x) - \alpha_0 \psi_0(x) - \alpha_1 \psi_1(x) - \cdots - \alpha_h \psi_h(x) \right]^2 = Min.$$

It is immediately clear that, in this way, we always get a strict least squares solution with respect to the distribution function $\omega(x)$ whatever the form of $g(x)$ may be.

In fact, the functions $f(x)$ and $g(x)$ are totally independent of one another, and, as is seen from (8), the distribution function $f(x)$ enters into the expansion of Tchebycheff for the regression function $g(x)$ only as a weight function which determines the weights to be allotted to the regression means in graduating the values of these by means of this series carried to a certain order, or, what is the same, by means of a parabola of the same order, the coefficients of which are determined according to the principle of least squares. Then it is clear that for practical purposes it is not necessary to derive the exact form of $f(x)$ in performing the expansion (6). The hypothetical distribution function $\omega(x)$ would be expected to give a satisfying result as soon as $\omega(x)$ in its main characteristics corresponds with the true distribution function $f(x)$.

I am going to work out the detailed solution for the following two usual forms of $\omega(x)$:

A/ Normal Error Function,

$$\omega(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$$

(22)

$$\xi = \frac{x - m_1}{\sigma_1},$$

B/ Pearson Type III Function,

$$\omega(\xi) = c\,(\xi + \alpha)^{\beta - 1} \cdot e^{-\alpha \xi}$$

(23)

$$\xi = \frac{x - m_1}{\sigma_1}; \quad \alpha = -\frac{1}{S}; \quad \beta = \alpha^2$$

S is the *skewness*, or

$$S = -\frac{\varepsilon_{30}}{2}. \tag{24}$$

In both cases the expressions for the terms of the series of Tchebycheff will be found to be very simple.

At first considering the polynomials $\psi_i(x)$ we are to have in accordance with (7), the distribution function being continuous,

$$\int_{-\infty}^{\infty} dx \cdot \omega(x) \cdot \psi_i(x) \cdot \psi_j(x) = 0 \qquad (i \neq j). \tag{25}$$

From this expression it may be concluded that the polynomials $\psi_i(x)$ are in case A/ the polynomials of Hermite, and in case B/ those of Laguerre. Both these kinds of polynomials are of well-known forms, and consequently the values of the e-coefficients as defined by (12) may easily be derived from propositions about these polynomials.

For the successive coefficients we have according to (9) the following expression

$$\overline{\alpha}_i = \frac{\int_{-\infty}^{\infty} dx \cdot \omega(x) \cdot \psi_i(x) \cdot g(x)}{\int_{-\infty}^{\infty} dx \cdot \omega(x) \cdot [\psi_i(x)]^2} . \tag{26}$$

Taking account of (13), (14), and (15) and introducing the notation

$$\overline{\nu}_{h1} = \int_{-\infty}^{\infty} dx \cdot \omega(x) \cdot x^h \cdot g(x) , \tag{27}$$

we obtain

$$\overline{\alpha}_i = \frac{\overline{\Delta}^{(i-1)}}{\overline{\Delta}^{(i)}} \left[\overline{\nu}_{i1} + \overline{e}_{i,i-1} \overline{\nu}_{i-1,1} + \cdots + \overline{e}_{i1} \overline{\nu}_{11} + \overline{e}_{i0} \overline{\nu}_{01} \right], \tag{28}$$

or, introducing the corresponding "standardized" moments $\overline{\varepsilon}_{ij}$ and putting

$$\overline{\varsigma}_{i0} = \overline{\varepsilon}_{i1} - \overline{\varepsilon}_{11} \overline{\varepsilon}_{i+1,0} \tag{29}$$

we get

$$\overline{\alpha}_i = \frac{\overline{\Delta}^{(i-1)}}{\overline{\Delta}^{(i)}} \left[\overline{\varsigma} + \overline{e}_{i:i-1} \overline{\varsigma}_{i-1,0} + \cdots + \overline{e}_{i2} \overline{\varsigma}_{30} \right] . \tag{30}$$

The coefficients $\frac{\overline{\Delta}^{(i-1)}}{\overline{\Delta}^{(i)}}$ and $\overline{e}_{ij}$ are determined by (14) and (15) when the moment values

(31)
$$\bar{\nu}_{h0} = \int_{-\infty}^{\infty} dx \cdot \omega(x) \cdot x^h,$$

respectively

(32)
$$\bar{\varepsilon}_{h0} = \int_{-\infty}^{\infty} d\xi \cdot \omega(\xi) \cdot \xi^h$$

are inserted in the determinants.

In the cases here considered we get very simple expressions for these determinants. We have, indeed, / W. Andersson, Op. cit., pp. 88 and 123 / in case of normal distribution

(33)
$$\frac{\bar{\Delta}^{(i-1)}}{\bar{\Delta}^{(i)}} = \frac{1}{i!},$$

and in case of Pearson type III distribution

(34)
$$\frac{\bar{\Delta}^{(i-1)}}{\bar{\Delta}^{(i)}} = \frac{1}{i! \prod_{h=1}^{i-1} (1+hS^2)}.$$

The values of the *e*-coefficients necessary for the computation of the terms of the series of Tchebycheff up to the fifth order are given in the following exposition.

Nor.	e_{ij}	Type III	Nor.	e_{ij}	Type III.
0	e_{10}	0	0	e_{43}	$12S$
0	e_{21}	$2S$	-6	e_{42}	$6(6S^2-1)$
-1	e_{20}	-1	0	e_{41}	$4S(6S^2-7)$
			3	e_{40}	$-3(6S^2-1)$
0	e_{32}	$6S$	0	e_{54}	$20S$
-3	e_{31}	$3(2S^2-1)$	-10	e_{53}	$10(12S^2-1)$
0	e_{30}	$-4S$	0	e_{52}	$20S(12S^2-5)$
			15	e_{51}	$5(24S^4-46S^2+3)$
			0	e_{50}	$-8S(12S^2-5)$

In order to derive the expressions for the computation of the moment quantities $\overline{\nu}_h$, or $\overline{\varepsilon}_h$, we denote the class-breadth by ϖ and the observed mean value of y in the p^{th} array of x by $\overline{y}_{x_p}$. The values of $\overline{\nu}_{h1}$ are then given by the following formula

$$(35) \qquad \overline{\nu}_{h1} = \sum_p I_{x_p} \cdot x_p^h \cdot \overline{y}_{x_p},$$

where

$$(36) \qquad I_{x_p} = \int_{x_p - \frac{\varpi}{2}}^{x_p + \frac{\varpi}{2}} dx \cdot \omega(x).$$

The computation is easily performed as soon as the function

$$(37) \qquad Q(x) = \int_{-\infty}^{x} dx \cdot \omega(x)$$

is known. In either case we have access to suitable tables of this function. For the Pearson type III function the "Tables for the incomplete Γ-function, edited by Karl Pearson" are to be used.

4. We will now make some general remarks concerning the relations between the different methods of computing regression parabolas touched upon in the preceding lines. We start with the general condition (8) for the determination of the coefficients:

$$\sum_x f(x) \cdot \left[g(x) - \alpha_0 \varphi_0(x) - \alpha_1 \varphi_1(x) - \cdots - \alpha_h \varphi_h(x)\right]^2 = Min.$$

It is seen that the expansion is determined by the marginal distribution function $f(x)$ and the regression function $g(x)$. If $f(x)$ and $g(x)$ are not the true functions of the population but the functions corresponding to the actual sample, the solution will give the sampling values of the coefficients. This is the solution advanced by Pearson, and consequently in his method no graduation of the data is performed in order to smooth out the influence of sampling irregularities on the values of the coefficients. Without any further considerations it is clear that methods which include an adjustment of the data in this respect are desirable. The problem is analogous with that occurring in the general theory of distributions. Among other facts of great importance that speak in favour of using mathematical functions for the description of distributions one is that we in this way are able to eliminate in some degree the accidental irregularities. When the regression is described by the series of Tchebycheff the smoothing process is evidently performed firstly by graduating the regression means by a parabola, and secondly by adjusting the parabola coefficients for the accidental irregularities. This latter adjustment has been accounted for by the two methods treated by the author. When using the rules of § 2 as principle for this adjustment the smoothing process is applied to the moment values involved in the general solution for the coefficients, and in the methods indicated in the preceding paragraphs we have used a weight function which is to be considered as a graduation of the observed marginal distribution of the independent variate.

As mentioned before we do not get a strict least squares solution when applying the rules of § 2. This is, however, of little practical importance, but it remains to see in what manner this solution is to be modified in order to become a least squares graduation of the observed array means.

When applying the rules of § 2 the product moments are computed from the following expression

$$\nu'_{h1} = \frac{1}{N} \sum_p n_{x_p} x_p^h \cdot \bar{y}_{x_p}, \tag{38}$$

where N is the total number of observations and n_{x_p} the number of observations in the p^{th} array of x. We suppose that the graduation is to be based on directly computed moment values up to the h^{th} order, h usually not being greater than six, in accordance with the rules of Thiele. The values of the marginal moments of the independent variate up to the h^{th} order indicate the distribution function $f(x, \nu'_{10}, \cdots, \nu'_{h0})$, which function is chosen as the theoretical distribution function determining the values of the marginal moments of orders above the h^{th}. A strict least squares solution with respect to the distribution function $f(x, \cdots, \nu'_{h0})$ may be worked out according to the formulas given in this memoir by taking $\omega_x = f(x, \nu'_{10}, \cdots, \nu'_{h0})$. In this case we have for the product moments the following values

$$\bar{\nu}'_{h1} = \sum_p I_{x_p} \cdot x_p^h \cdot \bar{y}_{x_p}, \tag{39}$$

where

$$I_{x_p} = \int_{x_p - \frac{\bar{\omega}}{2}}^{x_p + \frac{\bar{\omega}}{2}} dx \cdot f(x, \nu'_{10}, \cdots, \nu'_{h0}). \tag{40}$$

Subtracting (39) from (40) we get

$$\Delta\nu'_{h1} = \sum_p \left(I_{x_p} - \frac{n_{x_p}}{N} \right) \cdot x_p^h \cdot \bar{y}_{x_p}, \tag{41}$$

which consequently are the corrections to be added to the directly computed values of the product moments of the solution worked out in accordance with the rules of § 2, in order to obtain a strict least squares solution.

These corrections are easily computed as soon as the integrals I_{x_p} are determined. This task, however, would in some cases be somewhat arduous. If the general Pearson theory of frequency is applied we must sometimes resort to mechanical quadrature formulas.

a remark concerning the correction of the grouping of the moments $\overline{\nu}_{h1}$. According to the method of computing these characteristics we may regard them as mixed moments of a distribution having as its x-marginal distribution the function $\omega(x)$, the regression means being the observed ones. Thus we evidently can apply the usual methods of correcting computed moment values for the effect of grouping. By using the formulas of Sheppard we have to observe, however, that the moments involved in these formulas must be referred to the supposed semi-theoretical distribution.

5. *Numerical Illustrations.* In order to illustrate the application to observed data of the consideration above I have numerically treated a few populations—representative ones in that they are examples of correlation distributions of different degrees of skewness.

Example I. Case of slightly skew correlation. Pearson's example B. Example I:2 and II:2 of the cited memoir of the author. Population:Correlation between age and height of head in 2272 girls.

$$/x = \text{age};\ y = \text{height of head}\,/$$

$x_0 = 12.5$ yrs. $\quad \bar{x} - x_0 = +.2007 \quad \bar{\omega}_x = 1 \quad \xi = .3263\,x' - .0655$

$y_0 = 125.25$ m.m. $\quad \bar{y} - y_0 = -.6017 \quad \bar{\omega}_y = 2 \quad \eta = .2895\,y' + .1742.$

As regards the moment values I refer to the memoir of Pearson. These indicate that the marginal distribution of x may approximately be represented by the normal curve. Thus I take for ω_x the normal function.

We have to calculate the product moments $\overline{\nu}_{11}$, $\overline{\nu}_{21}$, and $\overline{\nu}_{31}$.

These computations may be performed by using the following scheme. The different values are derived from the correlation table given in Pearson's memoir.

The values of ξ correspond to the class ranges.

x'	$\bar{y}'$	$\frac{n_x}{N}$	ξ	I_x	$I_x - \frac{n_x}{N}$	$x' \cdot \bar{y}'_x$	$x'^2 \cdot \bar{y}'$	$x'^3 \cdot \bar{y}'$
- 9	- 5.000	.0004	- 3.165	.0015	.0011	45.000	- 405.000	3645.000
- 8	- 4.143	.0031	- 2.831	.0037	.0006	33.144	- 265.152	2121.216
- 7	- 3.889	.0079	- 2.513	.0084	.0005	27.223	- 190.561	1333.927
- 6	- 3.075	.0176	- 2.186	.0170	- .0006	18.450	- 110.700	664.200
- 5	- 2.474	.0335	- 1.860	.0311	- .0024	12.370	- 61.850	309.250
- 4	- 1.808	.0550	- 1.534	.0511	- .0039	7.232	- 28.928	115.712
- 3	- 1.763	.0779	- 1.208	.0756	- .0023	5.289	- 15.867	47.601
- 2	- 1.217	.1034	- 0.881	.1003	- .0031	2.434	- 4.868	9.736
- 1	- 1.054	.1149	- 0.555	.1199	.0050	1.054	- 1.054	1.054
0	- 0.680	.1360	- 0.229	.1296	- .0064	0.000	= 0.000	0.000
1	- 0.194	.1158	0.098	.1238	.0080	- 0.194	- 0.194	- 0.194
2	0.232	.0871	0.424	.1106	.0235	0.464	0.928	1.856
3	0.453	.0942	0.750	.0858	- .0084	1.359	4.077	12.231
4	0.642	.0713	1.077	.0605	- .0108	2.568	10.272	41.088
5	0.832	.0418	1.403	.0384	- .0034	4.160	20.800	104.000
6	0.885	.0268	1.729	.0218	- .0050	5.310	31.860	191.160
7	2.154	.0057	2.055	.0115	.0058	15.078	105.546	738.822
8	- 0.714	.0031	2.382	.0052	.0021	- 5.712	- 45.696	- 365.568
9	0.625	.0035	2.708	.0022	- .0013	5.625	50.625	455.625
10	0.000	.0009	3.034	.0008	- .0001	0.000	0.000	0.000

We get

$$\frac{\Sigma I_x x' \bar{y}'_x}{\Sigma I_x} = 2.98[illegible]9, \quad \frac{\Sigma I_x x'^2 \bar{y}'_x}{\Sigma I_x} = -6.6564, \quad \frac{\Sigma I_x x'^3 \bar{y}'_x}{\Sigma I_x} = 75.5197$$

and from these values

$$\bar{\nu}_{11} = 3.1123, \quad \bar{\nu}_{21} = -2.2376, \quad \bar{\nu}_{31} = 79.2110.$$

Sheppard's corrections for grouping have been applied.

For the corresponding standardized moments we obtain the following values:

$$\bar{\varepsilon}_{11} = 0.2941, \quad \bar{\varepsilon}_{21} = -0.0689, \quad \bar{\varepsilon}_{31} = 0.8000.$$

This leads to the following values of the $\bar{\varrho}$-coefficients defined by (29):

$$\bar{\varrho}_{30} = -0.0689, \quad \bar{\varrho}_{40} = -0.0823.$$

The values of the successive regression coefficients then become

$$\bar{\alpha}_1 = 0.2941, \quad \bar{\alpha}_2 = -0.0345, \quad \bar{\alpha}_3 = -0.0127.$$

Comparing these different values with the uncorrected ones we find

$$\begin{array}{llcl} \bar{\varepsilon}_{11} - r & = & 0.0000 & \qquad \bar{\varrho}_{30} - \varrho_{30} = +0.0021 \\ \bar{\varepsilon}_{21} - \varepsilon_{21} & = & -0.0086 & \qquad \bar{\varrho}_{40} - \varrho_{40} = -0.0337 \\ \bar{\varepsilon}_{31} - \varepsilon_{31} & = & +0.0511 & \qquad \bar{\alpha}_1 - \alpha_1 = +0.0000 \\ \bar{\varepsilon}_{30} - \varepsilon_{30} & = & -0.0365 & \qquad \bar{\alpha}_2 - \alpha_2 = +0.0010 \\ \bar{\varepsilon}_{40} - \varepsilon_{40} & = & +0.2894 & \qquad \bar{\alpha}_3 - \alpha_3 = -0.0056 \end{array}$$

We especially observe that the adjustments of the ϱ-coefficients are smaller than those of the moments of the same orders.

The adjusted coefficients result in the following regression parabola of the third order:

$$\bar{\eta}_\xi = +0.0345 + 0.3352\,\xi - 0.0345\,\xi^2 - 0.0137\,\xi^3.$$

The curve is drawn on diagram 1. For the sake of comparison the graph of the Pearson curve and that obtained by applying the rules of Thiele, the marginal being the normal curve, are given on the same diagram.

Example II. Case of moderately skew correlation. Population: Correlation between weight of newborn boy and weight of placenta; material supplied by the Maternity Hospital of Lund, Sweden. Example 2 in S. D. Wicksell: "Correlation Function of Type A, etc." /Kungl. Svenska Vetenskapsakademiens handlingar, Bd 58, Nr. 3/. N = 1223.

/x = weight of boy; y = weight of placenta/.

$x_0 = 3350$ gr. $\quad \bar{\omega}_x = 300 \quad \bar{x} - x_0 = +.4685 \quad \xi = .5940\,x' - .2783$

$y_0 = 630$ gr. $\quad \bar{\omega}_y = 80 \quad \bar{y} - y_0 = -.5715 \quad \eta = .6954\,y' + .3974$

The correlation table and the computed moment values are given in the said memoir of Wicksell. For ω_x we take the normal function.

Calculating the moments $\bar{\nu}_{11}$, etc. in the same manner as used in the first example we obtain the values,

$$\bar{\nu}_{11} = 1.5540, \quad \bar{\nu}_{21} = 0.3412, \quad \bar{\nu}_{31} = 11.7653$$

which give the following values of the standardized moments:

$$\bar{\varepsilon}_{11} = 0.6420, \quad \bar{\varepsilon}_{21} = 0.0837, \quad \bar{\varepsilon}_{31} = 1.7153.$$

The values are corrected for grouping.

We further get

$$\bar{\varrho}_{30} = +\,0.0837, \quad \bar{\varrho}_{40} = -\,0.2106$$

and

$$\bar{\tau}_1 = +\,0.6420, \quad \bar{\tau}_2 = +\,0.0419, \quad \bar{\tau}_3 = -\,0.0351.$$

The values of the adjustments of the different quantities are given below:

$\bar{\varepsilon}_{11} - r = -\,0.0035$	$\bar{\varrho}_{30} - \varrho_{30} = -\,0.0656$
$\bar{\varepsilon}_{21} - \varepsilon_{21} = +\,0.0196$	$\bar{\varrho}_{40} - \varrho_{40} = -\,0.0173$
$\bar{\varepsilon}_{31} - \varepsilon_{31} = -\,0.2132$	$\bar{\tau}_1 - \tau_1 = -\,0.0035$
$\bar{\varepsilon}_{30} - \varepsilon_{30} = +\,0.1320$	$\bar{\tau}_2 - \tau_2 = -\,0.0328$
$\bar{\varepsilon}_{40} - \varepsilon_{40} = -\,0.2870$	$\bar{\tau}_3 - \tau_3 = -\,0.0031.$

The correction of τ_2 is rather great, but not greater than was to be expected with consideration to the roughness of the fit of the hypothetical marginal distribution function. It is clear that when applying the solution of my previous paper in this example we should use a type IV curve for the marginal distribution. The unadjusted values of the parabola coefficients are also in this case easily computed, but the calculation of the adjustments by which the solution is turned into a least squares solution would be very laborious.

In order to illustrate the suitability of the several methods I have drawn the following curves on diagram 2: 1/ unadjusted solution, hypothetical marginal distribution being the normal function; 2/ adjusted solution, hypothetical marginal distribution being the normal function; 3/ unadjusted solution, marginal distribution being Pearson's type IV function.

The equation of the second curve is

$$\overline{\eta}_{\xi} = -0.0419 + 0.7473\,\xi + 0.0419\,\xi^2 - 0.0351\,\xi^3$$

The third curve is undoubtedly best fitted to the data.

Example III. Case of extremely skew correlation. The correlation between the age of bachelor and the age of spinster at marriage, Sweden 1911–1920. Example I:4 and II:7 in the cited memoir of the author. N = 321908.

$/x$ = age of spinster; y = age of bachelor/

$x_0 = 27.5$ yrs. $\quad \overline{\omega}_x = 5 \quad \bar{x} - x_0 = -.3131 \quad \xi = .9515\,x' + .2929$

$y_0 = 27.5$ yrs. $\quad \overline{\omega}_y = 5 \quad \bar{y} - y_0 = +.2824 \quad \eta = .8506\,y' - .2424.$

The moment values as given in the said memoir indicate that we can use Pearson's type III function as hypothetical distribution function for the x-marginal distribution. From the moment values as computed in the cited memoir we obtain the following values of the constants of this function:

$$\alpha = 1.4312 \qquad \beta = 2.0483\,.$$

It is to be remarked that for our purposes the computation of the constant C is not necessary.

For the e-coefficients we get the following values:

$e_{21} = -1.3974$	$e_{32} = -4.1922$	$e_{43} = -\ 8.3844$
$e_{20} = -1.0000$	$e_{31} = -0.0708$	$e_{42} = +11.5752$
	$e_{30} = +2.7948$	$e_{41} = +11.3771$
		$e_{40} = -\ 5.7876.$

For the unadjusted values of the ρ-coefficients and the successive regression coefficients we further get

$$\rho_{30} = +0.1787 \quad \rho_{40} = +0.5255 \quad \rho_{50} = +2.8763$$
$$\alpha_1 = +0.5535 \quad \alpha_2 = +0.05192$$
$$\alpha_3 = -0.0122669 \quad \alpha_4 = +0.003097.$$

Computing the corresponding adjusted values by use of "Tables of the incomplete Γ-function" we obtain

$$\bar{\rho}_{30} = +0.1723 \quad \bar{\rho}_{40} = +0.4638 \quad \bar{\rho}_{50} = +2.0851$$
$$\bar{\alpha}_1 = +0.5528 \quad \bar{\alpha}_2 = +0.05789$$
$$\bar{\alpha}_3 = -0.014647 \quad \bar{\alpha}_4 = +0.001097$$

Sheppard's corrections have been applied in both cases. The differences between the adjusted and the unadjusted values are

$\bar{\varepsilon}_{11} - r = -0.0007$	$\bar{\varepsilon}_{30} - \varepsilon_{30} = 0.0000$
$\bar{\varepsilon}_{21} - \varepsilon_{21} = -0.0034$	$\bar{\varepsilon}_{40} - \varepsilon_{40} = -0.4762$
$\bar{\varepsilon}_{31} - \varepsilon_{31} = -0.3237$	$\bar{\varepsilon}_{50} - \varepsilon_{50} = -2.0405$
$\bar{\varepsilon}_{41} - \varepsilon_{41} = -1.4548$	$\bar{\alpha}_1 - \alpha_1 = -0.0007$
$\bar{\rho}_{30} - \rho_{30} = -0.0064$	$\bar{\rho}_{40} - \rho_{40} = -0.0617$
$\bar{\rho}_{50} - \rho_{50} = -0.7912$	$\bar{\alpha}_2 - \alpha_2 = +0.00597$
	$\bar{\alpha}_3 - \alpha_3 = -0.001978$
	$\bar{\alpha}_4 - \alpha_4 = -0.002000.$

The parabolas of the third and the fourth orders are the following ones:

Unadjusted values of the coefficients:

$$\bar{\eta}_\xi = -.0873 + .4848\,\xi + .1050\,\xi^2 - .01267\,\xi^3$$
$$\bar{\eta}_\xi = -.1053 + .5171\,\xi + .1409\,\xi^2 - .03864\,\xi^3 + .003097\,\xi^4.$$

Adjusted values of the coefficients:

$$\bar{\eta}_\xi = -.0988 + .4729\,\xi + .1193\,\xi^2 - .01465\,\xi^3$$
$$\bar{\eta}_\xi = -.1052 + .4854\,\xi + .1320\,\xi^2 - .02384\,\xi^3 + .001097\,\xi^4.$$

The graphs are drawn on diagrams 3 and 4.

The results indicated by the few examples treated in this paragraph clearly point out that the Tchebycheff expansion cannot be considered as a least squares graduation of the observed regression means when the moment values involved in the solution are determined in accordance with the rules laid out in § 2. As regards the practical applicability of such a solution, however, this circumstance is of little importance, because the curve in this case is found to give as good, and sometimes a better representation of the regression than a strict least squares graduation. Further, as the calculation of the moments of the first few orders is often required for other purposes than the determination of the regression curve, the computation of the unadjusted solution in these cases is arithmetically very simple. Not having access to the moment values we may perhaps in some cases consider the direct computation of the adjusted solution as performed in example I to be the simplest method. The adjusting of correctly determined unadjusted solutions would certainly very seldom be of real gain.

Stockholm, September 1933.

S. D. Wicksell

Note on Dr. Andersson's Paper.

In an extensive memoir, *Researches into the theory of Regression,* Dr. W. Andersson has worked out a very simple and widely applicable numerical method of computing curved regressions. The general principle on which this method was founded Dr. Andersson has kindly attributed to me. It was laid out in my paper in the first number of the "Annals" Journal and may be stated as follows: After fitting a suitable univariate frequency function with a limited number of parameters—e.g. the normal curve or one of Pearson's types—to the marginal distribution of the independent variate, the moments of this function—which are all expressible in terms of the parameters—should be used in computing the regression coefficients, instead of the ordinary values (power means). Of course, when, in fitting the curve, the ordinary moments of lower orders have been used in determining the parameters, this procedure means that the moments of higher orders are theoretically expressed in terms of the moments of lower orders instead of being directly computed.

Applying this device to the ordinary least squares expressions for the regression coefficients, it was clear that a departure from the least square condition took place, but the chances were that this would not harm the result, and the computations would be much simplified. Dr. Andersson's investigation has shown that these expectations were highly justified.

During the official ventilation of the memoir, which was presented as Thesis for the degree of D.Ph., it was agreed that the method ought to be tested by a comparison with a theoretically very similar method in which the least squares condition was retained, although theoretical or semi-empirical weights were introduced instead of the purely empirical weights used in the method of Karl Pearson.

In the present paper Dr. Andersson has taken this question up and he shows that whereas the original (unadjusted) method is numerically simpler in application, it gives practically just as good regression curves as the new, adjusted method. In some cases he even considers the unadjusted solution to be the better one.

By this the incident may seem to be closed. I should, however, like to point out, in a few words, how very straightforward a principle it is, which lies behind this adjusted method.

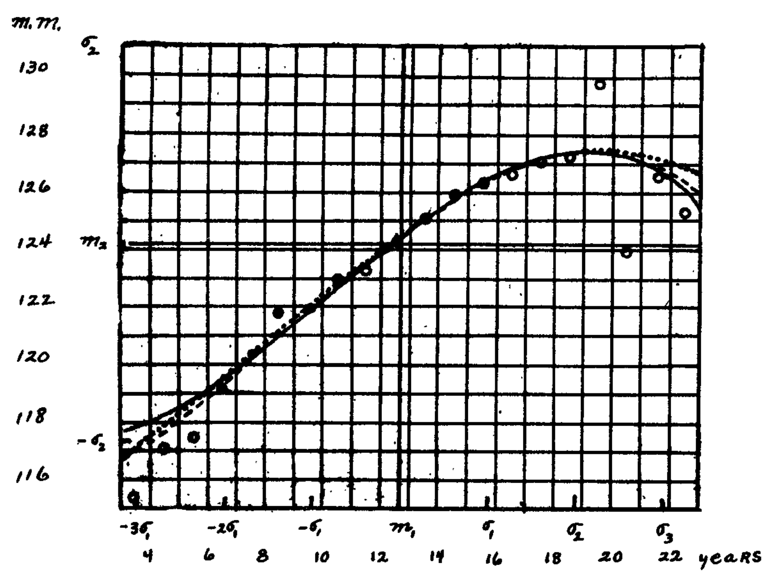
It is simply this: When a correlation table is given, the regression of y on x will not be affected by multiplying the frequencies within any x:array by a constant factor. Hence the following procedure will not affect the regression of y on x; i.e.. the process of reducing or adjusting the frequencies in the several x:arrays so that the marginal sums will be equal to the smoothed frequencies, corresponding to any mathematical curve which has been fitted to the marginal distribution. Thus, on applying Pearson's ordinary least squares solution to this adjusted table a least squares regression parabola would be obtained in which the marginal moments where those of the smoothed distribution, and also the mixed moments were, although only in a secondary degree, affected by the smoothing of the marginal. It is only in this last respect, i.e. as regards the mixed moments, that this method deviates from the one originally proposed.

In my opinion many curved regressions could be very easily and accurately enough computed by simply smoothing the marginal of the independent variate with a normal curve or, eventually, a Pearson Type III curve, and correspondingly adjusting the array frequencies. This method may work well even if the deviations of the actual distribution from the smoothed distribution are systematical.

Statistical Institute, University of Lund, November 1933.

Diagram 1

Cubics



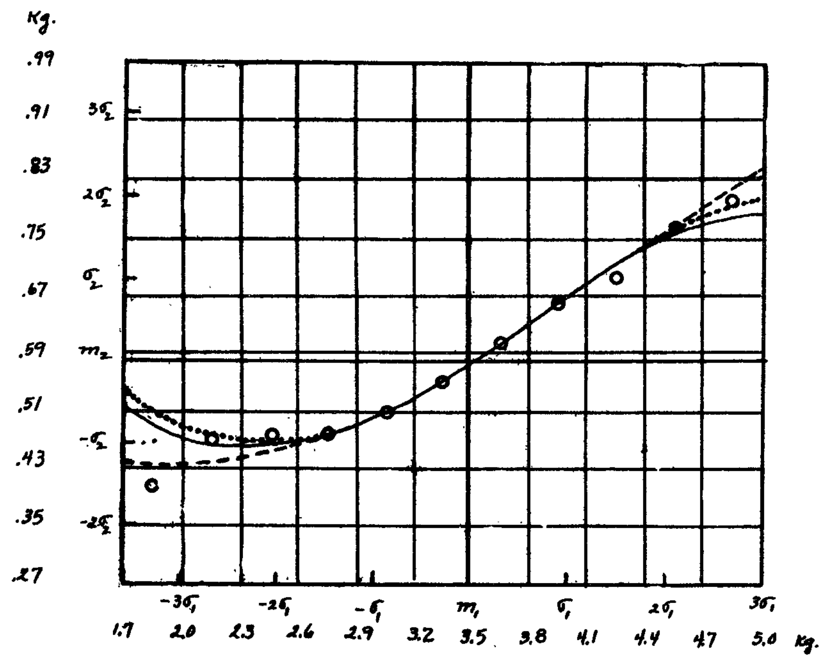
Unbroken curve: Adjusted solution, the hypothetical marginal distribution being the normal curve.

Dotted curve: Unadjusted solution, the hypothetical marginal distribution being the normal curve.

Dashed curve: Pearson's curve.

DIAGRAM 2

Cubics



Unbroken curve: Adjusted solution, the hypothetical marginal distribution being the normal curve.

Dotted curve: Unadjusted solution, the hypothetical marginal distribution being the normal curve.

Dashed curve: Unadjusted solution, the hypothetical marginal distribution being the Pearson Type IV curve.

DIAGRAM 3

Cubics

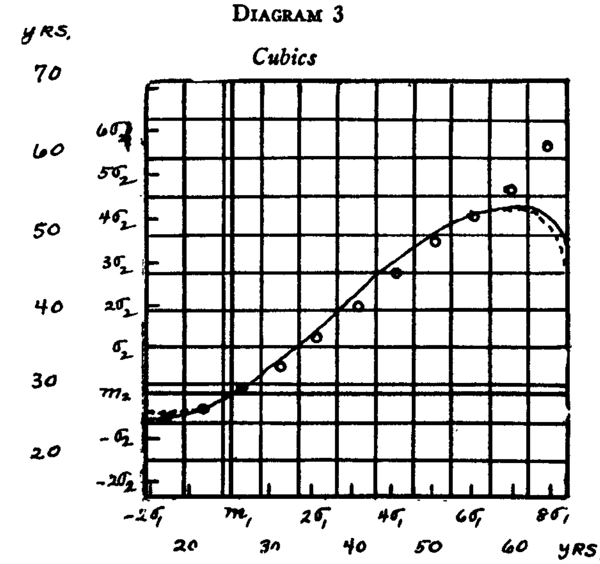
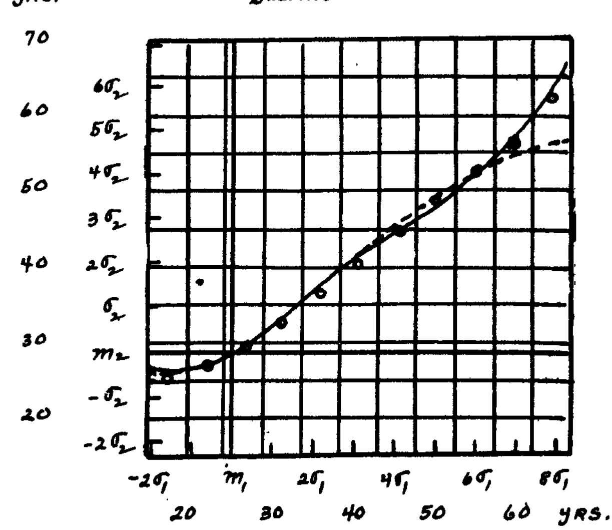


DIAGRAM 4

Quartics



Unbroken curves: Unadjusted solutions. Dashed curves: Adjusted solutions, the hypothetical marginal distribution being the Pearson Type III curve.