ON THE 'BEST' VALUES OF THE CONSTANTS IN FREQUENCY DISTRIBUTIONS.

By KIRSTINE SMITH.

(1) If we attempt to fit the normal or Gaussian curve to a system of observations, we almost invariably determine the constants $\bar{x}$ and σ of the equation

$$y = \frac{N}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x-\bar{x})^2}{\sigma^2}}$$

by *the method of moments*. This method of moments has been extended by Thiele, Pearson, Lipps and others to obtain the constants involved in various skew frequency curves and series. It is an undoubtedly utile and accurate method; but the question of whether it gives the 'best' values of the constants has not been very fully studied. It is perfectly true that if we deal with *individual* observations then the method of moments gives, with a somewhat arbitrary definition of what is to be a maximum, the 'best' values for σ and $\bar{x}$ in the above equation to the Gaussian. Pearson* has shown that the method of moments agrees with the method of least squares in the case where the distribution is given by a high order parabola, and accordingly the method of moments is likely to give a very good result, when an expansion by Maclaurin's Theorem would closely give a frequency function. But the method of least squares itself can now-a-days hardly be spoken of as more than a utile and accurate method of fit, indeed its utility, owing to the cumbersome nature of the equations which frequently arise, is often far less than that of the method of moments.

Gauss' original proof that the probability of the observed individual results was a maximum when $\bar{x}$ and σ have been determined by moments has led to the extension of the conception that for *grouped* data, and for other results than the Gaussian curve, the 'best' values of the constants must be given by the lowest possible moments. This is of course not true For example, if we had as frequency curve

$$y = y_0 e^{-\frac{1}{4}\frac{(x-\bar{x})^4}{\sigma^4}},$$

and used individual observations, then the Gaussian 'best' value for $\bar{x}$ would be that found by determining the point for which the third moment coefficient

* *Biometrika*, Vol. I. pp. 267–70.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

vanished, and the 'best' value of σ would be determined by $\sigma = \sqrt[4]{\mu_4}$, where μ_4 is to be taken about the point for which $\mu_3 = 0$*.

From another standpoint, however, the 'best values' of the frequency constants may be said to be those for which

$$\chi^2 = \frac{S(n_s - \bar{n}_s)^2}{\bar{n}_s}$$

is a minimum, where n_s is the observed frequency and $\bar{n}_s$ the theoretical frequency of the sth group†. For when χ^2 is a minimum then P, the probability of occurrence of a result as divergent as or more divergent than the observed, will be a maximum, or the frequency constants will have been so chosen as to make the probability P of results, as divergent from theory as the observed data occurring, a maximum.

It sounds somewhat paradoxical, but it is none the less true to say that the 'best value' of the mean is not necessarily the mean value, nor the 'best value' of the mean square deviation necessarily the mean square deviation‡. I shall illustrate this in the following five cases:

I. Fit of a normal curve to unilateral data.

II. Fit of a normal curve to bilateral data.

III. Fit of a Poisson limit to the binomial.

IV. Fit of a binomial to binomial data.

V. Fit of regression lines.

The general method is as follows. Suppose f to be any independent frequency constant; then χ^2 is to be a maximum with the variation of f. Accordingly we have from

$$1 + \chi^2 = S\left(\frac{n_s^2}{\bar{n}_s}\right)$$

* University of London, Honours B.Sc., *Papers in Statistics*, Thursday, Oct. 28, 1915.

† *Phil. Mag.* Vol. L. p. 157, 1900.

‡ There is a point of some philosophical interest here which deserves further consideration. As is well known the Gaussian demonstration depends on making the product

$$P\left\{\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\frac{(x_s-\bar{x})^2}{\sigma^2}}\right\},$$

s being taken so as to include each individual observation, a maximum by varying σ and $\bar{x}$, the result being that the 'best' values are found from the first two moments. Now it will be observed that this is not the same idea as lies in the χ^2 test of goodness of fit. The conception of 'goodness' in that case is that we should measure the probability of a drawing from a certain population giving as divergent or *a more divergent result than that observed*. In other words while the Gaussian test makes a *single ordinate* of a generalised frequency surface a maximum, the χ^2 test makes a real probability, namely the *whole volume* lying outside a certain contour surface defined by χ^2 a maximum. Logically this seems the more reasonable, for the above product used in the Gaussian proof is not a probability at all. To make it a probability it must be multiplied by the product $\{\delta x_s\}$, and then the probability of the actually observed result, namely $x_1, x_2, \ldots x_s, \ldots x_q$, will of course be infinitely small, and what is made a maximum is an infinitely small probability. The exact meaning of $P\{\delta x_s\}$ when x_s is an actual observation is obscure, but it appears that the probability for constant indefinitely small ranges of the variates *in the neighbourhood* of the observed values is made a maximum. But probability means the frequency of recurrence in a repeated series of trials and this probability is in the case supposed *indefinitely small*. It seems far more reasonable to make a finite probability, i.e. the probability of a divergence as great *or greater* than the observed a maximum, i.e. to use the χ^2 test and not the Gaussian principle.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

a number of equations of type

$$S\left(\frac{n_s^2}{\bar{n}_s^2}\frac{d\bar{n}_s}{df}\right)=0 \quad \ldots\ldots\ldots\ldots(1).$$

These equations will generally be far too involved to be directly solved. Accordingly we proceed thus: We suppose that the values of the frequency constants given by the method of moments are good starting-points, and we put, if $\bar{f}$ denote the moment value of a frequency constant, $f=\bar{f}+\Delta f$. Accordingly if there be a number $f_1, f_2, \ldots f_q$ of independent frequency constants, we shall have a series of equations to find $\Delta f_1, \Delta f_2, \ldots \Delta f_q$ of the type

$$\begin{aligned}0=S\left\{\frac{n_s^2}{\bar{n}_s^2}\left[\frac{d\bar{n}_s}{df_1}\right]\right\}&+S\left\{\frac{n_s^2}{\bar{n}_s^2}\left(\left[\frac{d^2\bar{n}_s}{df_1^2}\right]-\frac{2}{\bar{n}_s}\left[\left(\frac{d\bar{n}_s}{df_1}\right)^2\right]\right)\right\}\Delta f_1\\&+S\left\{\frac{n_s^2}{\bar{n}_s^2}\left(\left[\frac{d^2\bar{n}_s}{df_1 df_2}\right]-\frac{2}{\bar{n}_s}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_2}\right]\right)\right\}\Delta f_2\\&+\ldots\ldots\ldots\ldots\\&+S\left\{\frac{n_s^2}{\bar{n}_s^2}\left(\left[\frac{d^2\bar{n}_s}{df_1 df_q}\right]-\frac{2}{\bar{n}_s}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_q}\right]\right)\right\}\Delta f_q \quad \ldots\ldots(2\,a),\end{aligned}$$

where a square bracket round the differential coefficients signifies that the frequency constants $f_1, f_2, \ldots f_q$ therein are to be given their moment values $\bar{f}_1, \bar{f}_2, \ldots \bar{f}_q$. These values are of course also to be used in $\bar{n}_s$.

Since $S(\bar{n}_s)=N$, it is clear that

$$S\left(\frac{d\bar{n}_s}{df_1}\right)=S\left(\frac{d\bar{n}_s}{df_2}\right)=S\left(\frac{d^2\bar{n}_s}{df_1 df_2}\right)=\text{etc.}=0.$$

Accordingly the above equations may be reduced to the type

$$\begin{aligned}0=S\left\{\frac{n_s^2-\bar{n}_s^2}{\bar{n}_s^2}\left[\frac{d\bar{n}_s}{df_1}\right]\right\}&+S\left\{\frac{n_s^2-\bar{n}_s^2}{\bar{n}_s^2}\left[\frac{d^2\bar{n}_s}{df_1^2}\right]-\frac{2n_s^2}{\bar{n}_s^3}\left[\left(\frac{d\bar{n}_s}{df_1}\right)^2\right]\right\}\Delta f_1\\&+S\left\{\frac{n_s^2-\bar{n}_s^2}{\bar{n}_s^2}\left[\frac{d^2\bar{n}_s}{df_1 df_2}\right]-\frac{2n_s^2}{\bar{n}_s^3}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_2}\right]\right\}\Delta f_2\\&+\ldots\ldots\ldots\ldots\\&+S\left\{\frac{n_s^2-\bar{n}_s^2}{\bar{n}_s^2}\left[\frac{d^2\bar{n}_s}{df_1 df_q}\right]-\frac{2n_s^2}{\bar{n}_s^3}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_q}\right]\right\}\Delta f_q.\end{aligned}$$

It might reasonably be anticipated that terms involving the product of Δf and $(n_s^2-\bar{n}_s^2)/\bar{n}_s^2$ could be neglected in the first place and accordingly that we should have as approximate type

$$\begin{aligned}\tfrac{1}{2}S\left\{\frac{n_s^2-\bar{n}_s^2}{\bar{n}_s^2}\left[\frac{d\bar{n}_s}{df_1}\right]\right\}=S\left[\frac{n_s^2}{\bar{n}_s^3}\left(\frac{d\bar{n}_s}{df_1}\right)^2\right]\Delta f_1&+S\left\{\frac{n_s^2}{\bar{n}_s^3}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_2}\right]\right\}\Delta f_2\\&+\ldots+S\left\{\frac{n_s^2}{\bar{n}_s^3}\left[\frac{d\bar{n}_s}{df_1}\frac{d\bar{n}_s}{df_q}\right]\right\}\Delta f_q \quad \ldots\ldots\ldots(2\,b),\end{aligned}$$

but this approximation has not in every case numerically justified itself, and thus it cannot be invariably used as more than a reasonable starting-off point.

(2) *Fit of a Normal Curve.*

Differentiating $$\bar{n}_s=\frac{N}{\sqrt{2\pi}\sigma}\int_{x_s}^{x_{s+1}}e^{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}}dx,$$

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

and then putting

$$z_s = \frac{N}{\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x_s - m)^2}{\sigma^2}} \quad \text{and} \quad x_s/\sigma = h_s,$$

we have on substituting the differentials in $(2\,a)$:

$$\left.\begin{aligned}
0 = \bar{\sigma} S &\left\{\frac{n_s^2}{\bar{n}_s^2}[z_{s+1} - z_s]\right\} \\
&+ \Delta\bar{m}\left(S\left\{\frac{n_s^2}{\bar{n}_s^2}[h_{s+1}z_{s+1} - h_s z_s]\right\} + 2N\,S\left\{\frac{n_s^2}{\bar{n}_s^2}[z_{s+1} - z_s]^2\right\}\right) \\
&+ \Delta\bar{\sigma}\left(S\left\{\frac{n_s^2}{\bar{n}_s^2}[-z_{s+1} + z_s + h^2_{s+1}z_{s+1} - h_s^2 z_s]\right\} + 2N\,S\left\{\frac{n_s^2}{\bar{n}_s^2}[z_{s+1} - z_s][h_{s+1}z_{s+1} - h_s z_s]\right\}\right) \\
0 = \bar{\sigma} S &\left\{\frac{n_s^2}{\bar{n}_s^2}[h_{s+1}z_{s+1} - h_s z_s]\right\} \\
&+ \Delta\bar{m}\left(S\left\{\frac{n_s^2}{\bar{n}_s^2}[-z_{s+1} + z_s + h^2_{s+1}z_{s+1} - h_s^2 z_s]\right\} + 2N\,S\left\{\frac{n_s^2}{\bar{n}_s^2}[z_{s+1} - z_s][h_{s+1}z_{s+1} - h_s z_s]\right\}\right) \\
&+ \Delta\bar{\sigma}\left(S\left\{\frac{n_s^2}{\bar{n}_s^2}[(h^3_{s+1}z_{s+1} - h_s^3 z_s) - 2(h_{s+1}z_{s+1} - h_s z_s)]\right\} + 2N\,S\left\{\frac{n_s^2}{\bar{n}_s^2}[h_{s+1}z_{s+1} - h_s z_s]^2\right\}\right)
\end{aligned}\right\} \ldots (3),$$

the differential coefficients of χ^2 being

$$\left.\begin{aligned}
\frac{d(\chi^2)}{dm} &= \frac{N}{\bar{\sigma}} S\left\{\frac{n_s^2}{\bar{n}_s^2}(z_{s+1} - z_s)\right\} \\
\text{and} \qquad \frac{d(\chi^2)}{d\sigma} &= \frac{N}{\bar{\sigma}} S\left\{\frac{n_s^2}{\bar{n}_s^2}(h_{s+1}z_{s+1} - h_s z_s)\right\}
\end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (4).$$

Illustration I. Fit of Normal Curve to Unilateral Data.

Our first illustration treats a series of measurements by Bradley discussed by Bessel*. The mean of the observations is fixed, for in dealing with the observations Bessel has added positive and negative variations together.

TABLE I. *Measurements of Right Ascension.*

Limits	Observed	Gaussian curve by moments	Gaussian curve improved by minimum χ^2
0″·0—0″·1	114	101·61	98·63
0″·1—0″·2	84	84·12	82·59
0″·2—0″·3	53	57·65	57·91
0″·3—0″·4	24	32·71	34·00
0″·4—0″·5	14	15·36	16·72
0″·5—0″·6	6	5·974	6·881
0″·6—0″·7	3	1·923	2·372
0″·7—0″·8	1	·5122	·6843
0″·8—0″·9	1	·1370	·2053

* Emanuel Czuber, *Theorie der Beobachtungsfehler*, p. 192. Search has been made in vain in the *Fundamenta Astronomiae* for the original data in order to remove the unilateral limitation.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

$\bar{\sigma}$ was found equal to 2·282542 and the second formula of (3) gave the value 2·341735 for σ. As $\Delta\sigma$ was found so large that the approximation could not be expected to be very good, the following values of $\frac{d(\chi^2)}{d\sigma}$ were calculated from the second formula of (4):

σ	$\frac{d(\chi^2)}{d\sigma}$
2·282542	− 32·53
$\frac{1}{\cdot 43}$ = 2·325581	− 11·81
$\frac{1}{\cdot 42}$ = 2·380952	+ 8·06

By interpolation in this table σ = 2·355860 was found as the value for which $\frac{d(\chi^2)}{d\sigma}$ equals zero, and this is the σ of the improved Gaussian given above.

From χ^2 the 'goodness of fit' P was found:

	χ^2	P	$\frac{d(\chi^2)}{d\sigma}$
Gaussian ...	10·833	0·211	− 32·53
Impr. Gaussian	9·720	0·285	+ 0·20

As will be seen the better fit is obtained by making σ bigger than the Gaussian value, the improvement therefore cannot be looked upon as a correction for grouping. On the contrary the Sheppard correction would have given σ = 2·264214 and have raised χ^2 to 11·52. Thus we see that although the two methods give close values for P, the 'better value' is obtained as it should be from the lesser value of $d(\chi^2)/d\sigma$.

(3) *Illustration II. Fit of a Normal Curve to Bilateral Data.*

For the next illustration I have used a table giving frequencies of cephalic index in Bavarian skulls*. Both σ and m have here been varied. As the formulae (3) are somewhat laborious to work with, the approximations were used roughly suggested by the process on p. 264, but the results were not satisfactory† and these approximate results are therefore not given here. But $\frac{d(\chi^2)}{dm}$ and $\frac{d(\chi^2)}{d\sigma}$ for the two

* J. Ranke, *Beiträge zur physischen Antropologie der Baiern*, München, 1883. The table includes the material from Tables I–VI and VIII–X inclusive which may be treated as typically 'Alt-Baierisch.'

† In fact the calculation of the exact value of $\frac{d^2(\chi^2)}{d\sigma^2}$ showed that the part of it neglected in formula (2 *b*) was about $\frac{1}{17}$ of the whole value. It essentially arose from the one tail group, this being [illegible] of the whole neglected part. As $\frac{n_s^2 - \bar{n}_s^2}{\bar{n}_s^2}$ for this group was only as big as 1·0348, the approximate formula (2 *b*) cannot be expected to be of great value for the normal curve.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

Gaussians found in this way were used for interpolation purposes and their constants are therefore given in the following table under (*b*) and (*c*). By (*a*) is indicated the Gaussian from which we started, namely that found by moments, Sheppard's correction being used.

Assuming $\frac{d(\chi^2)}{dm}$ and $\frac{d(\chi^2)}{d\sigma}$ to be linear functions of σ and m, we determined from the cases (*a*), (*b*) and (*c*) values of σ and m, given under (*d*), so as to make the differential coefficients zero. In the same way we found at last from the cases (*a*), (*b*) and (*d*) the constants of the Gaussian (*e*), the constants of which will be found in the following table. As will be seen we have succeeded in bringing the values of $\frac{d(\chi^2)}{dm}$ and $\frac{d(\chi^2)}{d\sigma}$ near to zero, certainly close enough for all practical purposes.

TABLE II.

	m	σ	χ^2	P	$\frac{d(\chi^2)}{dm}$	$\frac{d(\chi^2)}{d\sigma}$
(*a*)	83·06889	3·431833	10·205	·895	− ·57	+ 14·42
(*b*)	83·01498	3·358380	10·301	·891	− 10·56	− 9·97
(*c*)	82·98832	3·331365	11·048	·854	− 15·89	− 20·76
(*d*)	83·05329	3·349421	10·108	·899	− 4·59	− 12·10
(*e*)	83·07774	3·385991	9·858	·909	+ ·07	+ ·71

TABLE III.

	Observed	Gaussian curve by moments	Gaussian improved by minimum χ^2
75 and under	9·5	12·3387	11·3504
76	12·5	12·6842	12·0767
77	17	22·0702	21·3463
78	37	35·2942	34·6005
79	55	51·8794	51·4323
80	71·5	70·0925	70·1100
81	82	87·0421	87·6432
82	116	99·3519	100·4734
83	98	104·2329	105·6275
84	107	100·5128	101·8352
85	82	89·0879	90·0352
86	74	72·5781	72·9998
87	58	54·3468	54·2778
88	34·5	37·4049	37·0099
89	19	23·6625	23·1422
90	10	13·7588	13·2703
91	8	7·3532	6·9782
92 and over	9	6·3093	5·7910

(4) *Fit of a Poisson Limit to the Binomial*. For a Poisson limit with the general term $\frac{e^{-\bar{m}}\bar{m}^s}{s!}$ we find

$$\frac{d(\chi^2)}{dm} = S\left(\frac{n_s^2}{\bar{n}_s}\frac{\bar{m}-s}{\bar{m}}\right) \quad\ldots\ldots(5),$$

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

and putting $m = \bar{m} + \Delta m$,

$$\Delta m = \frac{S\left(\frac{n_s^2}{\bar{n}_s}(s-\bar{m})\right)}{S\left(\frac{n_s^2}{\bar{n}_s}((s-\bar{m})^2+s)\right)}\,\bar{m} \quad \text{..........................(6).}$$

Starting with $\bar{m}$ equal to the mean of the observations I have found the improved values in the following two numerical examples.

Illustration III. The first table given by L. Whitaker* contains the number of deaths per day of women over 85 years, published in the *Times* newspaper during the years 1910–1912.

TABLE IV.

Number of deaths per day	Observed	Poisson by first moment	Poisson improved by minimum χ^2
0	364	336·250	331·133
1	376	397·302	396·334
2	218	234·720	237·186
3	89	92·446	94·630
4	33	27·308	28·316
5	13	6·4532	6·7782
6	2	1·2708	1·3521
7	1	0·2508	0·2715

The m, χ^2, P and $\frac{d(\chi^2)}{dm}$ calculated from (5) were determined for the two distributions as given in Table V.

TABLE V.

	m	χ^2	P	$\frac{d(\chi^2)}{dm}$
Poisson	1·181569	15·226	·0332	− 35·61
Poisson improved	1·196903	14·943	·0361	− 0·75

Illustration IV. As our second illustration we have taken a table of phagocytic frequencies published by Major McKendrick†.

TABLE VI.

No. of Deaths	Observed	Poisson by first moment	Poisson improved by minimum χ^2
0	620	605·924	600·676
1	282	303·568	306·164
2	79	76·044	78·026
3	16	12·699	13·257
4	2	1·5906	1·6892
5	1	·1738	·1881

* *Biometrika*, Vol. x. p. 67.

† *Proceedings of the London Mathematical Society*, Vol. xiii. 1913, p. 401.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

The numerical values of the constants of the series and of the 'goodness of fit' are

TABLE VII.

	m	χ^2	P	$\frac{d(\chi^2)}{dm}$
Poisson	·501000	6·865	·231	− 41·86
Poisson improved	·509700	6·672	·246	− 1·21

This table is of interest because it illustrates the apparent paradox, already seen in the case of the second Gaussian curve illustration, that the 'mean' is not necessarily the 'best value' of the constant termed the 'mean.'

(5) *Fit of a Binomial to Binomial Data.*

Let $\bar{n}_s$ be equal to the $(s+1)$th term of the binomial $(p+q)^l$, where $p+q=1$, or to

$$p^{n-s}(1-p)^s\frac{l(l-1)\dots(l-s+1)}{s!};$$

we then find

$$\frac{dn_s}{dp}=\bar{n}_s\frac{l-pl-s}{p(1-p)}=\bar{n}_s\frac{m-s}{p(1-p)},$$

where m is the mean or stand for $l(1-p)$,

$$\frac{d^2n_s}{dp^2}=\frac{n_s}{p^2(1-p)^2}\{(l-pl-s)^2-(l-pl-s)(1-2p)-lp(1-p)\}$$

$$=\frac{n_s}{p^2(1-p^2)}\{(m-s)^2+(m-s)(1-2p)+mp\},$$

$$\frac{dn_s}{dl}=\bar{n}_s\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right),$$

$$\frac{d^2n_s}{dl^2}=\bar{n}_s\left\{\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)^2-\frac{1}{l^2}-\frac{1}{(l-1)^2}-\dots-\frac{1}{(l-s+1)^2}\right\},$$

$$\frac{d^2n_s}{dl\,dp}=\frac{\bar{n}_s}{p(1-p)}\left\{(m-s)\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)+(1-p)\right\}.$$

Hence we have

$$\left.\begin{aligned}\frac{d(\chi^2)}{dp}&=-S\left[\frac{n_s^2}{\bar{n}_s}\frac{m-s}{p(1-p)}\right]\\ \frac{d(\chi^2)}{dl}&=-S\left[\frac{n_s^2}{\bar{n}_s}\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)\right]\end{aligned}\right\}\dots\dots\dots\dots(7),$$

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

and the equations (2 *a*) take the form

$$\left.\begin{aligned}
S\left(\frac{n_s^2}{\bar{n}_s}\,\frac{m-s}{p(1-p)}\right) &= S\left(\frac{n_s^2}{\bar{n}_s p^2(1-p)^2}[(m-s)^2+(m-s)(1-2p)+mp]\right)\Delta p \\
&\quad + S\left(\frac{n_s^2}{\bar{n}_s p(1-p)}\left[(m-s)\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)-(1-p)\right]\right)\Delta l \\
S\left(\frac{n_s^2}{\bar{n}_s}\left[\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right]\right) & \\
&= S\left(\frac{n_s^2}{\bar{n}_s p(1-p)}\left[(m-s)\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)-(1-p)\right]\right)\Delta p \\
&\quad + S\left(\frac{n_s^2}{\bar{n}_s}\left[\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)^2+\frac{1}{l^2}+\frac{1}{(l+1)^2}+\dots+\frac{1}{(l-s+1)^2}\right]\right)\Delta l
\end{aligned}\right\}\ (8),$$

while the approximate formulae of the type (2 *b*) are

$$\left.\begin{aligned}
S\left(\frac{n_s^2}{\bar{n}_s}\,\frac{m-s}{p(1-p)}\right) &= 2S\left(\frac{n_s^2}{\bar{n}_s p^2(1-p)^2}[m-s]^2\right)\Delta p \\
&\quad + 2S\left(\frac{n_s^2}{\bar{n}_s p(1-p)}\left[(m-s)\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)\right]\right)\Delta l \\
S\left(\frac{n_s^2}{\bar{n}_s}\left[\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right]\right) & \\
&= 2S\left(\frac{n_s^2}{\bar{n}_s p(1-p)}\left[(m-s)\left(\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right)\right]\right)\Delta p \\
&\quad + 2S\left(\frac{n_s^2}{\bar{n}_s}\left[\log_e p+\frac{1}{l}+\frac{1}{l-1}+\dots+\frac{1}{l-s+1}\right]\right)\Delta l
\end{aligned}\right\}\ (9).$$

Illustration V. Weldon's Dice Data

For illustration are used the following data due to the late Professor W. F. R. Weldon*. They give the observed frequencies of dice with five or six points when a throw of twelve dice was made 26306 times.

TABLE VIII.

Number of dice in cast with 5 or 6 points	Observed frequency	Binomial by method of moments	Improved binomial (*a*) by χ^2 a minimum	Improved binomial (*b*) by χ^2 a minimum
0	185	189·679	190·651	190·659
1	1149	1154·441	1157·607	1157·600
2	3265	3223·426	3226·085	3225·959
3	5475	5461·01	5458·07	5457·78
4	6114	6253·64	6245·98	6245·71
5	5194	5101·31	5095·82	5095·79
6	3067	3041·04	3041·47	3041·69
7	1331	1335·82	1339·55	1339·81
8	403	429·627	432·815	432·984
9	105	98·865	100·351	100·419
10	14	15·5133	15·9413	15·9595
11	4	1·57640	1·65879	1·66210

* *Phil. Mag.* July, 1900, p. 167.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

Fitting the frequencies from the end by means of two moments we obtain the binomial

$$(6658208 + 3341792)^{12\cdot 126379},$$

the terms of which are given in the table above under the head Binomial.

From these starting values of p and l we found by the equations (8) the constants of the improved binomial (*a*) $p = \cdot 6674922$ and $l = 12\cdot 188945$.

A comparison between the coefficients of the two sets of formulae (8) and (9) gave the result that they only diverged by between 1·4 and 5 per mille of their value. As $\frac{n_s^2 - \bar{n}_s^2}{\bar{n}_s^2}$ for the tail group was as big as 5·44, we are from this justified in expecting the approximate formulae (9) to be useful for binomial data.

Starting from the improved binomial (*a*) another improved binomial (*b*) was found by means of the formulae (9). As will be seen I only succeeded in diminishing $\frac{d(\chi^2)}{dp}$ by raising $\frac{d(\chi^2)}{dl}$, and χ^2 came out with exactly the same value as by the former formula. The constants for the improved binomial (*b*) are $p = \cdot 6675432$ and $l = 12\cdot 191141$.

The constants illustrating the 'goodness of fit' were found as follows:

TABLE IX.

	χ^2	P	$\frac{d(\chi^2)}{dp}$	$\frac{d(\chi^2)}{dl}$
Binomial	11·643	·390	159·47	− 8·15
Improved Binomial (*a*)	11·513	·401	26·02	− ·84
" (*b*)	11·513	·401	− ·02	+ 1·96

It will be seen from the above illustrations that the probability of happening as determined by the χ^2 test of 'goodness of fit' being a maximum can always be made somewhat greater than the same probability deduced from a fit by the method of moments, which at any rate for the Gaussian curve is usually assumed to be the 'best.'

(6) *On the 'Best' Values of the Constants of Regression Curves.*

If we apply the test of 'goodness of fit' to regression curves as recently indicated by Pearson* modifying Slutsky's methods†, we shall experience the same divergence between the curves of regression found by the method of least squares and the curves calculated so as to make χ^2 a minimum, as we found when dealing with frequency distributions.

In the paper cited χ^2 for a regression curve is given as

$$\chi^2 = S\left(\frac{n_p(m_p - \bar{m}_p)^2}{\sigma^2_{\bar{n}_p}}\right) \quad \text{..............................(10)},$$

* *Biometrika*, Vol. XI. pp. 239 *et seq.*

† *Journal of the Royal Statistical Society*, Vol. LXXVII. pp. 78–84.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

where m is the mean of the pth array of the sample of size M from a population of size N, while $\bar{m}_p$ is the theoretical mean as found from the regression curve, $n_p, = \bar{n}_p M/N$, is the mean frequency and $\sigma_{\bar{n}_p}$ the mean standard deviation of the pth array in the samples. The difficulty in applying the 'goodness of fit' test lies in finding adequate values for n_p and $\sigma_{\bar{n}_p}$. Let us assume them to be found. The 'best' values of the constants $f_1, f_2, \ldots$ of the regression curve, i.e. the values which make χ^2 a minimum, will then be found from equations of the type

$$0 = -2S\left\{\frac{n_p}{\sigma^2_{\bar{n}_p}}(m_p - \bar{m}_p)\frac{d\bar{m}_p}{df_1}\right\} + S\left\{(m_p - \bar{m}_p)^2 \frac{d\left(\frac{n_p}{\sigma^2_{\bar{n}_p}}\right)}{df_1}\right\} \quad \ldots\ldots(11).$$

As will be seen these equations fall into the equations resulting from using the method of least squares if $\frac{n_p}{\sigma^2_{\bar{n}_p}}$ is independent of the constants of the regression curve and at the same time for the different arrays proportional to the n_p of the sample. Even if our sample be derived from truly Gaussian data, these conditions will only approximately be satisfied, the $\sigma_{\bar{n}_p}$'s, although constant, being dependent upon the constants of the regression curve and the n_p of the formula not being really the sample value.

Supposing $\frac{n_p}{\sigma^2_{\bar{n}_p}}$ to be independent of the constants of the regression line $\bar{m}_p = ax + b$, the equations (11) take the form

$$S\{v_p(m_p - ax - b)x\} = 0,$$
$$S\{v_p(m_p - ax - b)\} = 0,$$

when we put v_p for

$$\frac{\frac{n_p}{\sigma^2_{\bar{n}_p}}}{S\left(\frac{n_p}{\sigma^2_{\bar{n}_p}}\right)}$$

From these equations we find

$$a = \frac{S(v_p m_p x) \,.\, S(v_p) - S(v_p m_p)\, S(x)}{S(v_p x^2) \,.\, S(v_p) - \{S(v_p x)\}^2}$$

and

$$b = \frac{S(v_p m_p)}{S(v_p)} - a\frac{S(v_p x)}{S(v_p)},$$

formulae agreeing with those derived from the method of least squares if v_p equals the marginal frequencies of the sample. But not agreeing with them if, for example, the material be heteroscedastic.

(7) *Illustration VI. Auricular Height of School Girls.*

This example was first used by Pearson in the memoir on skew correlation* and later as an illustration of the test of 'goodness of fit' of regression curves†.

* *Drapers' Company Research Memoirs*, Biometric Series II. p. 34.

† *Biometrika*, Vol. XI. p. 253.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

For the present use theoretical values of n_p and $\sigma^2_{\bar{x}_p}$ were determined, from which the values of v_p given in Table X are calculated. The n_p and v_p of the table represent the weights given to the means of arrays respectively by the method of least squares and by our method of making χ^2 a minimum. It will be seen that our method throws

TABLE X.

Age	n_p observed	v_p	m_p observed	m_p from χ^2 a minimum	m_p from least squares
3—4	1	5·3790	115·25	117·76	117·95
4—5	7	13·7170	116·96	118·44	118·61
5—6	18	28·5973	117·47	119·13	119·27
6—7	40	56·0527	119·10	119·81	119·94
7—8	76	95·3828	120·30	120·49	120·60
8—9	125	146·023	121·63	121·17	121·26
9—10	177	199·783	121·72	121·86	121·92
10—11	235	243·414	122·82	122·54	122·59
11—12	261	271·704	123·14	123·22	123·25
12—13	309	277·232	123·89	123·90	123·91
13—14	263	259·386	124·86	124·59	124·58
14—15	198	223·505	125·71	125·27	125·24
15—16	214	172·851	126·16	125·95	125·90
16—17	162	121·965	126·53	126·63	126·57
17—18	95	75·7303	126·91	127·32	127·23
18—19	61	43·0926	127·02	128·00	127·89
19—20	13	21·2448	129·56	128·68	128·55
20—21	7	8·09110	123·82	129·36	129·22
21—22	8	6·42326	126·50	130·05	129·88
22—23	2	2·42653	125·25	130·73	130·54

the weight more to the first half part of the groups of ages than the method of least squares. This is due to the heteroscedasticity of the material, the $\sigma^2_{\bar{x}_p}$ varying from 27·2776 in the youngest group to 60·4676 in the eldest. The two last columns of Table X contain the m_p calculated from our regression formula and from the usual formula; as might be expected our m_p's are closer to the means of the observations for the younger groups of ages and differ more for the higher ages than do the m_p values obtained by the method of least squares. The χ^2 calculated by (10) are for the two cases 18·45 and 18·67 and we have only raised the 'goodness of fit' P from ·543 to ·558 although the weighting in the two methods appeared sensibly different.

The usual regression line is

$$m_p = 124{\cdot}0467 + {\cdot}662979\,(x_p - 12{\cdot}7007),$$

124·0467 and 12·7007 being the general means, and regression line from the χ^2 formula may be written

$$m_p = 124{\cdot}0411 + {\cdot}682455\,(x_p - 12{\cdot}7007)$$

from which is seen that it passes not far from the mean.

In a similar way I have treated the regression of ages on height of head. Also I have here calculated the heteroscedasticity and have had to use a parabola to

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

Diagram I. Comparison of Regression Straight Lines found by method of Least Squares and by χ^2 Test.

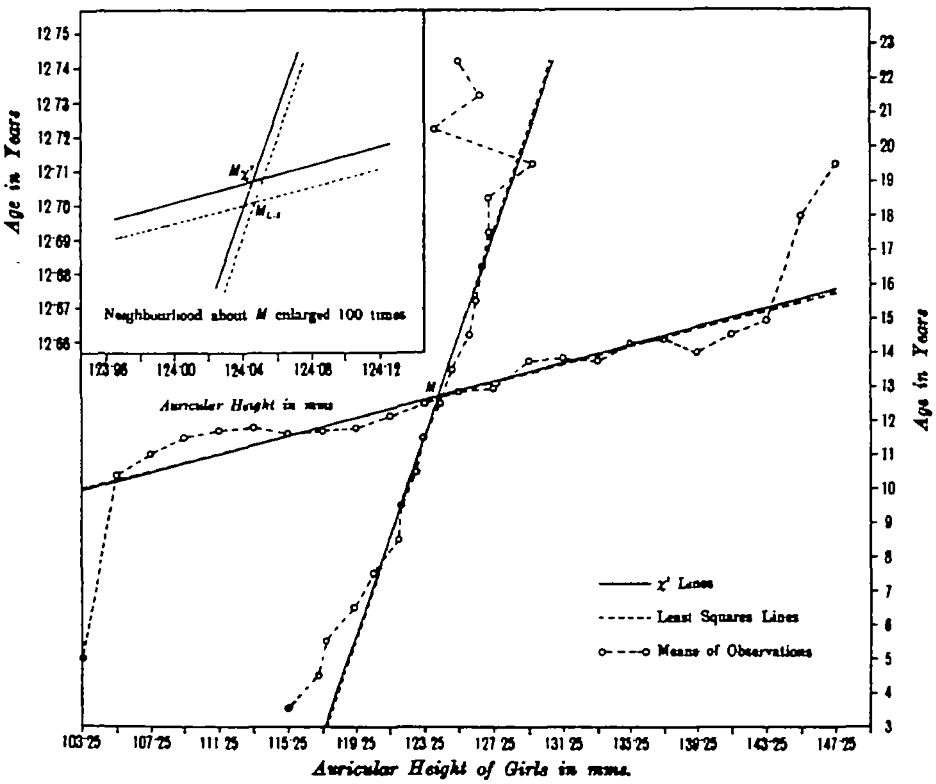
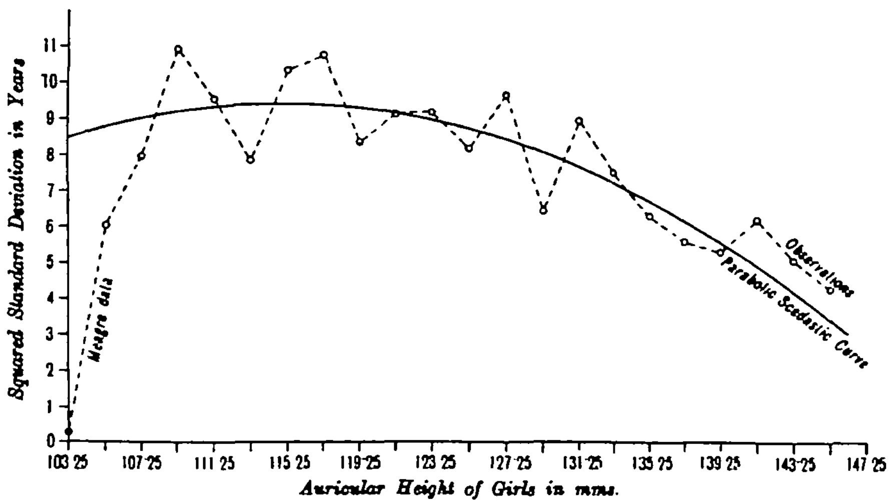


Diagram II. Scedastic Curve of Age on Auricular Height in Girls.



Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

represent $\sigma^2_{\bar{n}_p}$, the squared standard deviation of the arrays of same height, to obtain a reasonable description; this is shown on the diagram. The marginal frequencies of the height variate could be expressed fairly well by a Gaussian curve. These theoretical values of $\sigma^2_{\bar{n}_p'}$ and n_p' are given in Table XI together with the weights

$$v' = \frac{\dfrac{n_p'}{\sigma^2_{\bar{n}_p'}}}{S\left(\dfrac{n_p}{\sigma^2_{\bar{n}_p'}}\right)}$$

calculated from them.

TABLE XI.

Millims	$\sigma^2_{\bar{n}_p'}$ theoretical	n_p' theoretical	v'	n_p' observed	m_p' observed	m_p' from χ^2 minimum	m_p' from least squares
102·25—104·25	8·456	4·73	4·7123	2	5·00	9·92	9·99
104·25—106·25	8·748	6·62	6·3809	10	10·40	10·18	10·25
106·25—108·25	8·987	13·89	13·0282	10	11·10	10·45	10·51
108·25—110·25	9·172	26·80	24·6339	27	11·54	10·72	10·77
110·25—112·25	9·304	47·59	43·1217	56	11·71	10·99	11·03
112·25—114·25	9·382	77·76	69·8648	59	11·81	11·26	11·29
114·25—116·25	9·408	116·90	104·750	115	11·62	11·53	11·55
116·25—118·25	9·380	161·71	145·332	142	11·70	11·80	11·81
118·25—120·25	9·298	205·83	186·597	244	11·80	12·06	12·08
120·25—122·25	9·164	241·06	221·744	265	12·15	12·33	12·34
122·25—124·25	8·976	259·78	243·960	281	12·52	12·60	12·60
124·25—126·25	8·735	257·59	248·580	265	12·83	12·87	12·86
126·25—128·25	8·441	235·02	234·710	219	12·98	13·14	13·12
128·25—130·25	8·093	197·30	205·508	197	13·78	13·41	13·38
130·25—132·25	7·692	152·41	167·023	131	13·85	13·67	13·64
132·25—134·25	7·238	108·33	126·167	88	13·78	13·94	13·90
134·25—136·25	6·730	70·85	88·7361	77	14·28	14·21	14·16
136·25—138·25	6·170	42·64	58·2529	52	14·40	14·48	14·42
138·25—140·25	5·556	23·61	35·8204	20	14·05	14·75	14·69
140·25—142·25	4·888	12·03	20·7416	16	14·56	15·02	14·95
142·25—144·25	4·168	5·64	11·4040	11	14·95	15·29	15·21
144·25—146·25	3·394	2·43	6·0407	4	18·00	15·55	15·47
146·25—148·25	2·567	1·49	4·8835	1	19·50	15·82	15·73

The usual regression line is

$$m_p' = 12{\cdot}7007 + {\cdot}130489\,(y_p - 124{\cdot}0467),$$

and the line for which χ^2 is a minimum is

$$m_p' = 12{\cdot}7071 + {\cdot}1342345\,(y_p - 124{\cdot}0467).$$

For χ^2 were found in the two cases the values 44·411 and 44·109 and for the 'goodness of fit' P the values ·0047 and ·0051*.

* A case was purposely chosen in which the regression was known to be far from linear, in order to ascertain whether this fact itself would separate at all widely the least square and χ^2 regression lines.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020

The intersection point of the two χ^2 regression lines is $m = 124 \cdot 0453$, $m' = 12 \cdot 7070$, which is seen to be very near to the general means. Introducing that point into the equations of the lines, they take the form

$$m_p' = 12 \cdot 7070 + \cdot 1342345\,(y_p - 124 \cdot 0453),$$

$$m_p = 124 \cdot 0453 + \cdot 682455\,(x_p - 12 \cdot 7070).$$

From the slopes of the lines we find the value ·3027 for the correlation coefficient, whereas the method of least squares gives the value ·2941.

Although we have found the material to be decidedly heteroscedastic and the weighting of the two series of means rather different from that of the marginal frequencies, we nevertheless see that the resulting regression lines differ very little from the ordinary regression lines, both the deviations of the means and the correlation coefficient derived from them being less than their probable errors.

(8) The conclusions to be drawn from the present investigation are:

(i) The definition of 'best,' which leads to the method of moments being considered 'best' and incidentally to the method of least squares being considered 'best,' is undoubtedly somewhat arbitrary. If we use Pearson's 'Goodness of Fit' test, then the method of moments is not necessarily the 'best,' the best value of the constant termed the mean is not necessarily the mean, nor generally the best value of the correlation coefficient between two variates that calculated by the moments and product moment method.

(ii) On the other hand the present numerical illustrations appear to indicate that but little practical advantage is gained by a great deal of additional labour, the values of P are only slightly raised—probably always within their range of probable error. In other words the investigation justifies the method of moments as giving excellent values of the constants with nearly the maximum value of P or it justifies the use of the method of moments, if the definition of 'best' by which that method is reached must at least be considered somewhat arbitrary.

The present paper was worked out in the Biometric Laboratory and I have to thank Professor Pearson for his aid throughout the work.

Downloaded from https://academic.oup.com/biomet/article/11/3/262/192813 by McGill University Libraries user on 04 October 2020