VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

ANY experiment may be regarded as forming an individual of a "population" of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

Now any series of experiments is only of value in so far as it enables us to form a judgment as to the statistical constants of the population to which the experiments belong. In a great number of cases the question finally turns on the value of a mean, either directly, or as the mean difference between the two quantities.

If the number of experiments be very large, we may have precise information as to the value of the mean, but if our sample be small, we have two sources of uncertainty:—(1) owing to the "error of random sampling" the mean of our series of experiments deviates more or less widely from the mean of the population, and (2) the sample is not sufficiently large to determine what is the law of distribution of individuals. It is usual, however, to assume a normal distribution, because, in a very large number of cases, this gives an approximation so close that a small sample will give no real information as to the manner in which the population deviates from normality: since some law of distribution must be assumed it is better to work with a curve whose area and ordinates are tabled, and whose properties are well known. This assumption is accordingly made in the present paper, so that its conclusions are not strictly applicable to populations known not to be normally distributed; yet it appears probable that the deviation from normality must be very extreme to lead to serious error. We are concerned here solely with the first of these two sources of uncertainty.

The usual method of determining the probability that the mean of the population lies within a given distance of the mean of the sample, is to assume a normal distribution about the mean of the sample with a standard deviation equal to $s/\sqrt{n}$, where s is the standard deviation of the sample, and to use the tables of the probability integral.

But, as we decrease the number of experiments, the value of the standard deviation found from the sample of experiments becomes itself subject to an increasing error, until judgments reached in this way may become altogether misleading.

In routine work there are two ways of dealing with this difficulty: (1) an experiment may be repeated many times, until such a long series is obtained that the standard deviation is determined once and for all with sufficient accuracy. This value can then be used for subsequent shorter series of similar experiments. (2) Where experiments are done in duplicate in the natural course of the work, the mean square of the difference between corresponding pairs is equal to the standard deviation of the population multiplied by $\sqrt{2}$. We can thus combine together several series of experiments for the purpose of determining the standard deviation. Owing however to secular change, the value obtained is nearly always too low, successive experiments being positively correlated.

There are other experiments, however, which cannot easily be repeated very often; in such cases it is sometimes necessary to judge of the certainty of the results from a very small sample, which itself affords the only indication of the variability. Some chemical, many biological, and most agricultural and large scale experiments belong to this class, which has hitherto been almost outside the range of statistical enquiry.

Again, although it is well known that the method of using the normal curve is only trustworthy when the sample is "large," no one has yet told us very clearly where the limit between "large" and "small" samples is to be drawn.

The aim of the present paper is to determine the point at which we may use the tables of the probability integral in judging of the significance of the mean of a series of experiments, and to furnish alternative tables for use when the number of experiments is too few.

The paper is divided into the following nine sections:

I. The equation is determined of the curve which represents the frequency distribution of standard deviations of samples drawn from a normal population.

II. There is shown to be no kind of correlation between the mean and the standard deviation of such a sample.

III. The equation is determined of the curve representing the frequency distribution of a quantity z, which is obtained by dividing the distance between the mean of a sample and the mean of the population by the standard deviation of the sample.

IV. The curve found in I. is discussed.

V. The curve found in III. is discussed.

VI. The two curves are compared with some actual distributions.

VII. Tables of the curves found in III. are given for samples of different size.

VIII and IX. The tables are explained and some instances are given of their use.

X. Conclusions.

Now 50 to 1 corresponds to three times the probable error in the normal curve and for most purposes would be considered significant; for this reason I have only tabled my curves for values of n not greater than 10, but have given the $n=9$ and $n=10$ tables to one further place of decimals. They can be used as foundations for finding values for larger samples*.

The table for $n=2$ can be readily constructed by looking out $\theta=\tan^{-1}z$ in Chambers' Tables and then $\cdot 5+\theta/\pi$ gives the corresponding value.

Similarly $\frac{1}{2}\sin\theta+\cdot 5$ gives the values when $n=3$.

There are two points of interest in the $n=2$ curve. Here s is equal to half the distance between the two observations. $\tan^{-1}\frac{s}{s}=\frac{\pi}{4}$ so that between $+s$ and $-s$ lies $2\times\frac{\pi}{4}\times\frac{1}{\pi}$ or half the probability, i.e. if two observations have been made and we have no other information, it is an even chance that the mean of the (normal) population will lie between them. On the other hand the second moment coefficient is

$$\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}}\tan^2\theta\,d\theta=\frac{1}{\pi}\Big[\tan\theta-\theta\Big]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=\infty,$$

or the standard deviation is infinite while the probable error is finite.

SECTION VI. *Practical Test of the foregoing Equations.*

Before I had succeeded in solving my problem analytically, I had endeavoured to do so empirically. The material used was a correlation table containing the height and left middle finger measurements of 3000 criminals, from a paper by W. R. Macdonell (*Biometrika*, Vol. I. p. 219). The measurements were written out on 3000 pieces of cardboard, which were then very thoroughly shuffled and drawn at random. As each card was drawn its numbers were written down in a book which thus contains the measurements of 3000 criminals in a random order. Finally each consecutive set of 4 was taken as a sample—750 in all—and the mean, standard deviation, and correlation† of each sample determined. The difference between the mean of each sample and the mean of the population was then divided by the standard deviation of the sample, giving us the z of Section III.

This provides us with two sets of 750 standard deviations and two sets of 750 z's on which to test the theoretical results arrived at. The height and left middle finger correlation table was chosen because the distribution of both was approximately normal and the correlation was fairly high. Both frequency curves, however, deviate slightly from normality, the constants being for height $\beta_1=\cdot 0026$, $\beta_2=3\cdot 175$, and for left middle finger lengths $\beta_1=\cdot 0030$, $\beta_2=3\cdot 140$, and in consequence there is a tendency for a certain number of larger standard deviations to occur than if the distributions were normal. This, however, appears to make very little difference to the distribution of z.

* E.g. if $n=11$, to the corresponding value for $n=9$, we add $\frac{7}{8}\times\frac{5}{6}\times\frac{3}{4}\times\frac{1}{2}\times\frac{1}{2}\cos^8\theta\sin\theta$: if $n=13$ we add as well $\frac{9}{10}\times\frac{7}{8}\times\frac{5}{6}\times\frac{3}{4}\times\frac{1}{2}\times\frac{1}{2}\cos^{10}\theta\sin\theta$ and so on.

† I hope to publish the results of the correlation work shortly.

Another thing which interferes with the comparison is the comparatively large groups in which the observations occur. The heights are arranged in 1 inch groups, the standard deviation being only 2·54 inches: while the finger lengths were originally grouped in millimetres, but unfortunately I did not at the time see the importance of having a smaller unit, and condensed them into two millimetre groups, in terms of which the standard deviation is 2·74.

Several curious results follow from taking samples of 4 from material disposed in such wide groups. The following points may be noticed:

(1) The means only occur as multiples of ·25.

(2) The standard deviations occur as the square roots of the following types of numbers n, $n + \cdot 19$, $n + \cdot 25$, $n + \cdot 50$, $n + \cdot 69$, $2n + \cdot 75$.

(3) A standard deviation belonging to one of these groups can only be associated with a mean of a particular kind; thus a standard deviation of $\sqrt{2}$ can only occur if the mean differs by a whole number from the group we take as origin, while $\sqrt{1 \cdot 69}$ will only occur when the mean is at $n \pm \cdot 25$.

(4) All the four individuals of the sample will occasionally come from the same group, giving a zero value for the standard deviation. Now this leads to an infinite value of z and is clearly due to too wide a grouping, for although two men may have the same height when measured by inches, yet the finer the measurements the more seldom will they be identical, till finally the chance that four men will have *exactly* the same height is infinitely small. If we had smaller grouping the zero values of the standard deviation might be expected to increase, and a similar consideration will show that the smaller values of the standard deviation would also be likely to increase, such as ·436, when 3 fall in one group and 1 in an adjacent group, or ·50 when 2 fall in two adjacent groups. On the other hand when the individuals of the sample lie far apart, the argument of Sheppard's correction will apply, the real value of the standard deviation being more likely to be smaller than that found owing to the frequency in any group being greater on the side nearer the mode.

These two effects of grouping will tend to neutralise each other in their effect on the mean value of the standard deviation, but both will increase the variability.

Accordingly we find that the mean value of the standard deviation is quite close to that calculated, while in each case the variability is sensibly greater. The fit of the curve is not good, both for this reason and because the frequency is not evenly distributed owing to effects (2) and (3) of grouping. On the other hand the fit of the curve giving the frequency of z is very good and as that is the only practical point the comparison may be considered satisfactory.

The following are the figures for height:—

Mean value of standard deviations;	calculated	2·027 ± ·021
" " "	observed	2·026
	Difference =	− ·001

Standard deviation of standard deviations:—

Calculated ·8556 ± ·015
Observed ·9066

Difference = + ·0510

Comparison of Fit. Theoretical Equation: $y = \frac{16 \times 750}{\sqrt{2\pi}\sigma^3} x^2 e^{-\frac{2x^2}{\sigma^2}}$.

Scale in terms of standard deviation of population	0 to ·1	·1 to ·2	·2 to ·3	·3 to ·4	·4 to ·5	·5 to ·6	·6 to ·7	·7 to ·8	·8 to ·9	·9 to 1·0	1·0 to 1·1	1·1 to 1·2	1·2 to 1·3	1·3 to 1·4	1·4 to 1·5	1·5 to 1·6	1·6 to 1·7	Greater than 1·7
Calculated frequency	1½	10½	27	45½	64½	78½	87	88	81½	71	58	45	33	23	15	9½	5½	7
Observed frequency	3	14½	24½	37½	107	67	73	77	77½	64	52½	49½	35	28	12½	9	11½	7
Difference	+1½	+4	−2½	−8	+42½	−11½	−14	−11	−4	−7	−5½	+4½	+2	+5	−2½	−½	+6	0

whence $\chi^2 = 48{\cdot}06$, $P = {\cdot}000{,}06$ (about).

In tabling the observed frequency, values between ·0125 and ·0875 were included in one group, while between ·0875 and ·0125 they were divided over the two groups. As an instance of the irregularity due to grouping I may mention that there were 31 cases of standard deviations 1·30 (in terms of the grouping) which is ·5117 in terms of the standard deviation of the population, and they were therefore divided over the groups ·4 to ·5 and ·5 to ·6. Had they all been counted in groups ·5 to ·6 χ^2 would have fallen to 29·85 and P would have risen to ·03. The χ^2 test presupposes *random* sampling from a frequency following the given law, but this we have not got owing to the interference of the grouping.

When, however, we test the z's where the grouping has not had so much effect we find a close correspondence between the theory and the actual result.

There were three cases of infinite values of z which, for the reasons given above, were given the next largest values which occurred, namely + 6 or − 6. The rest were divided into groups of ·1; ·04, ·05 and ·06, being divided between the two groups on either side.

The calculated value for the standard deviation of the frequency curve was 1 (± ·017) while the observed was 1·039. The value of the standard deviation is really infinite, as the fourth moment coefficient is infinite, but as we have arbitrarily limited the infinite cases we may take as an approximation $\frac{1}{\sqrt{1500}}$ from which the value of the probable error given above is obtained. The fit of the curve is as follows:—

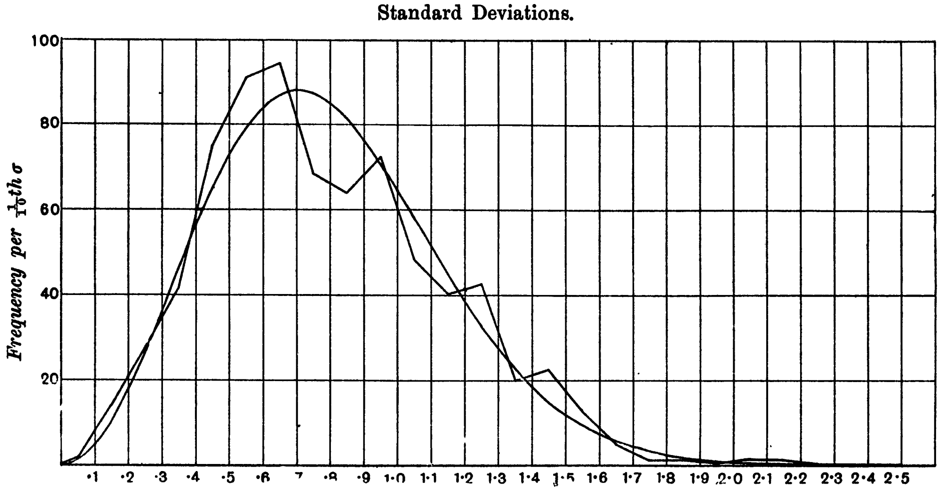
Comparison of Fit. Theoretical Equation: $y = \frac{2}{\pi}\cos^4\theta,\ z = \tan\theta.$

Scale of z	less than −3·05	−3·05 to −2·05	−2·05 to −1·55	−1·55 to −1·05	−1·05 to −·75	−·75 to −·45	−·45 to −·15	−·15 to +·15	+·15 to +·45	+·45 to +·75	+·75 to +1·05	+1·05 to +1·55	+1·55 to +2·05	+2·05 to +3·05	more than +3·05
Calculated frequency	5	9½	13½	34½	44½	78½	119	141	119	78½	44½	34½	13½	9½	5
Observed frequency	9	14½	11½	33	43½	70½	119½	151½	122	67½	49	26½	16	10	6
Difference	+4	+5	−2	−1½	−1	−8	+½	+10½	+3	−11	+4½	−8	+2½	+½	+1

whence $\chi^2 = 12{\cdot}44$, $P = {\cdot}56$.

This is very satisfactory, especially when we consider that as a rule observations are tested against curves fitted from the mean and one or more other moments of the observations, so that considerable correspondence is only to be expected; while this curve is exposed to the full errors of random sampling, its constants having been calculated quite apart from the observations.

Diagram III. Comparison of Calculated Standard Deviation Frequency Curve with 750 actual Standard Deviations.



The left middle finger samples show much the same features as those of the height, but as the grouping is not so large compared to the variability the curves fit the observations more closely. Diagrams III.* and IV. give the standard deviations and the z's for this set of samples. The results are as follows:—

* There are three small mistakes in plotting the observed values in Diagram III., which make the fit appear worse than it really is.

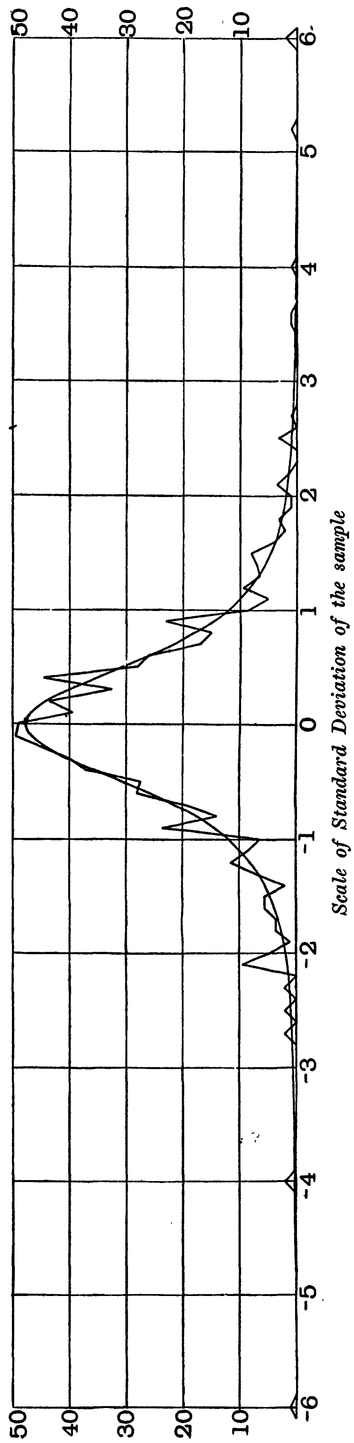


DIAGRAM IV. Comparison of the theoretical frequency curve $y=\frac{1500}{\pi}\left(1+\frac{x^2}{s^2}\right)^{-2}$, with an actual sample of 750 cases.

Mean value of standard deviations; calculated 2·186 ± ·023
" " " observed 2·179
Difference = − ·007

Standard deviation of standard deviations:—

Calculated ·9224 ± ·016
Observed ·9802
Difference = + ·0578

Comparison of Fit. Theoretical Equation: $y = \frac{16 \times 750}{\sqrt{2\pi}\sigma^3} x^2 e^{-\frac{2x^2}{\sigma^2}}$.

Scale in terms of standard deviation of population	0 to ·1	·1 to ·2	·2 to ·3	·3 to ·4	·4 to ·5	·5 to ·6	·6 to ·7	·7 to ·8	·8 to ·9	·9 to 1·0	1·0 to 1·1	1·1 to 1·2	1·2 to 1·3	1·3 to 1·4	1·4 to 1·5	1·5 to 1·6	1·6 to 1·7	greater than 1·7
Calculated frequency	1½	10½	27	45½	64½	78½	87	88	81½	71	58	45	33	23	15	9½	5½	7
Observed frequency	2	14	27½	51	64½	91	94½	68½	65½	73	48½	40½	42½	20	22½	12	5	7½
Difference	+½	+3½	+½	+5½	—	+12½	+7½	−19½	−16	+2	−9½	−4½	+9½	−3	+7½	+2½	−½	+½

whence $\chi^2 = 21·80$, $P = ·19$.

Calculated value of standard deviation 1 (± ·017)
Observed " " " ·982
Difference = − ·018

Comparison of Fit. Theoretical Equation: $y = \frac{2}{\pi} \cos^4 \theta$, $z = \tan \theta$.

Scale of z	less than −3·05	−3·05 to −2·05	−2·05 to −1·55	−1·55 to −1·05	−1·05 to −·75	−·75 to −·45	−·45 to −·15	−·15 to +·15	+·15 to +·45	+·45 to +·75	+·75 to +1·05	+1·05 to +1·55	+1·55 to +2·05	+2·05 to +3·05	more than +3·05
Calculated frequency	5	9½	13½	34½	44½	78½	119	141	119	78½	44½	34½	13½	9½	5
Observed frequency	4	15½	18	33½	44	75	122	138	120½	71	46½	36	11	9	6
Difference	−1	+6	+4½	−1	−½	−3½	+3	−3	+1½	−7½	+2	+1½	−2½	−½	+1

whence $\chi^2 = 7·39$, $P = ·92$.

A very close fit.

We see then that if the distribution is approximately normal our theory gives us a satisfactory measure of the certainty to be derived from a small sample in both the cases we have tested; but we have an indication that a fine grouping is of advantage. If the distribution is not normal, the mean and the standard deviation of a sample will be positively correlated, so that although both will have greater variability, yet they will tend to counteract each other, a mean deviating largely from the general mean tending to be divided by a larger standard deviation Consequently I believe that the tables at the end of the present paper may be used in estimating the degree of certainty arrived at by the mean of a few experiments, in the case of most laboratory or biological work where the distributions are as a rule of a 'cocked hat' type and so sufficiently nearly normal.

SECTION VII. *Tables of* $\frac{n-2}{n-3}\,\frac{n-4}{n-5}\cdots\begin{pmatrix}\frac{3}{2}\cdot\frac{1}{2}\, n \text{ odd}\\ \frac{2}{1}\cdot\frac{1}{\pi}\, n \text{ even}\end{pmatrix}\int_{-\frac{\pi}{2}}^{\tan^{-1}z}\cos^{n-2}\theta\,d\theta$

for values of n from 4 *to* 10 *inclusive.*

Together with $\frac{\sqrt{7}}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{7x^2}{2}}\,dx$ *for comparison when* $n = 10$.

$z\left(=\frac{x}{s}\right)$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	For comparison $\left(\frac{\sqrt{7}}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{7x^2}{2}}\,dx\right)$
·1	·5633	·5745	·5841	·5928	·6006	·60787	·61462	·60411
·2	·6241	·6458	·6634	·6798	·6936	·70705	·71846	·70159
·3	·6804	·7096	·7340	·7549	·7733	·78961	·80423	·78641
·4	·7309	·7657	·7939	·8175	·8376	·85465	·86970	·85520
·5	·7749	·8131	·8428	·8667	·8863	·90251	·91609	·90691
·6	·8125	·8518	·8813	·9040	·9218	·93600	·94732	·94375
·7	·8440	·8830	·9109	·9314	·9468	·95851	·96747	·96799
·8	·8701	·9076	·9332	·9512	·9640	·97328	·98007	·98253
·9	·8915	·9269	·9498	·9652	·9756	·98279	·98780	·99137
1·0	·9092	·9419	·9622	·9751	·9834	·98890	·99252	·99820
1·1	·9236	·9537	·9714	·9821	·9887	·99280	·99539	·99926
1·2	·9354	·9628	·9782	·9870	·9922	·99528	·99713	·99971
1·3	·9451	·9700	·9832	·9905	·9946	·99688	·99819	·99986
1·4	·9531	·9756	·9870	·9930	·9962	·99791	·99885	·99989
1·5	·9598	·9800	·9899	·9948	·9973	·99859	·99926	·99999
1·6	·9653	·9836	·9920	·9961	·9981	·99903	·99951	
1·7	·9699	·9864	·9937	·9970	·9986	·99933	·99968	
1·8	·9737	·9886	·9950	·9977	·9990	·99953	·99978	
1·9	·9770	·9904	·9959	·9983	·9992	·99967	·99985	
2·0	·9797	·9919	·9967	·9986	·9994	·99976	·99990	
2·1	·9821	·9931	·9973	·9989	·9996	·99983	·99993	
2·2	·9841	·9941	·9978	·9992	·9997	·99987	·99995	
2·3	·9858	·9950	·9982	·9993	·9998	·99991	·99996	
2·4	·9873	·9957	·9985	·9995	·9998	·99993	·99997	
2·5	·9886	·9963	·9987	·9996	·9998	·99995	·99998	
2·6	·9898	·9967	·9989	·9996	·9999	·99996	·99999	
2·7	·9908	·9972	·9991	·9997	·9999	·99997	·99999	
2·8	·9916	·9975	·9992	·9998	·9999	·99998	·99999	
2·9	·9924	·9978	·9993	·9998	·9999	·99998	·99999	
3·0	·9931	·9981	·9994	·9998	—	·99999	—	—