THE PROBABLE ERROR OF A MEAN

By STUDENT

Introduction

Any experiment may he regarded as forming an individual of a "population" of experiments which might he 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 greater 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 he 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 load 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 he 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 he 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 call 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 inquiry.

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.

Section 1

Samples of n individuals are drawn out of a population distributed normally, to find an equation which shall represent the frequency of the standard deviations of these samples.

If s be the standard deviation found from a sample $x_1 x_2 \dots x_n$ (all these being measured from the mean of the population), then

$$s^2 = \frac{S(x_1^2)}{n} - \left(\frac{S(x_1)}{n}\right)^2 = \frac{S(x_1^2)}{n} - \frac{S(x_1^2)}{n^2} - \frac{2S(x_1 x_2)}{n^2}.$$

Summing for all samples and dividing by the number of samples we get the moan value of s^2, which we will write $\bar{s}^2$:

$$\bar{s}^2 = \frac{n\mu_2}{n} - \frac{n\mu_2}{n^2} = \frac{\mu_2(n-1)}{n},$$

where μ_2 is the second moment coefficient in the original normal distribution of x: since x_1, x_2, etc. are not correlated and the distribution is normal, products involving odd powers of x_1 vanish on summing, so that $\frac{2S(x_1 x_2)}{n^2}$ is equal to 0.

If M'_R represent the Rth moment coefficient of the distribution of s^2 about the end of the range where $s^2 = 0$,

$$M'_1 = \mu_2 \frac{(n-1)}{n}.$$

Again

$$\begin{aligned}
s^4 &= \left\{ \frac{S(x_1^2)}{n} - \left(\frac{S(x_1)}{n}\right) \right\}^2 \\
&= \left(\frac{S(x_1^2)}{n}\right)^2 - \frac{2S(x_1^2)}{n}\left(\frac{S(x_1)}{n}\right)^2 + \left(\frac{S(x_1)}{n}\right)^4 \\
&= \frac{S(x_1^4)}{n^2} + \frac{2S(x_1^2 x_2^2)}{n^2} - \frac{2S(X_1^4)}{n^3} - \frac{4S(x_1^2 x_2^2)}{n^3} + \frac{S(x_1^4)}{n^4} \\
&+ \frac{6S(x_1^2 x_2^2)}{n^4} + \text{other terms involving odd powers of } x_1, \text{ etc. which} \\
&\qquad \text{will vanish on summation.}
\end{aligned}$$

Now $S(x_1^4)$ has n terms, but$S(x_1^2 x_2^2)$ has $\frac{1}{2}n(n-1)$, hence summing for all samples and dividing by the number of samples, we get

$$\begin{aligned}
M'_2 &= \frac{\mu_4}{n} + \mu_2^2 \frac{(n-1)}{n} - \frac{2\mu_4}{n^2} - 2\mu_2^2 \frac{(n-1)}{n^2} + \frac{\mu_4}{n^3} + 3\mu_2^2 \frac{(n-1)}{n^3} \\
&= \mu_4 n^3 \{n^2 - 2n + 1\} + \frac{\mu_2^2}{n^3}(n-1)\{n^2 - 2n + 3\}.
\end{aligned}$$

Now since the distribution of x is normal, $\mu_4 = 3\mu_2^2$, hence

$$M'_2 = \mu_2^2 \frac{(n-1)}{n^3}\{3n - 3 + n^2 - 2n + 3\} = \mu_2^2 \frac{(n-1)(n+1)}{n^2}.$$

In a similar tedious way I find

$$M'_3 = \mu_2^3 \frac{(n-1)(n+1)(n+3)}{n^3}$$

and

$$M'_4 = \mu_2^4 \frac{(n-1)(n+1)(n+3)(n+5)}{n^4}.$$

The law of formation of these moment coefficients appears to be a simple one, but I have not seen my way to a general proof.

If now M_R be the Rth moment coefficient of s^2 about its mean, we have

$$M_2 = \mu_2^2 \frac{(n-1)}{n^3}\{(n+1)-(n-1)\} = 2\mu_2^2 \frac{(n-1)}{n^2}.$$

$$M_3 = \mu_2^3 \left\{ \frac{(n-1)(n+1)(n+3)}{n^3} - \frac{3(n-1)}{n}.\frac{(2(n-1)}{n^2} - \frac{(n-1)^3}{n^3} \right\}$$

$$= \mu_2^3 \frac{(n-1)}{n^3}\{n^2+4n+3-6n+6-n^2+2n-1\} = 8\mu_2^3 \frac{(n-1)}{n^3},$$

$$M_4 = \frac{\mu_2^4}{n^4}\left\{(n-1)(n+1)(n+3)(n+5) - 32(n-1)^2 - 12(n-1)^3 - (n-1)^4\right\}$$

$$= \frac{\mu_2^4(n-1)}{n^4}\{n^3+9n^2+23n+15-32n+32$$

$$-12n^2+24n-12-n^3+3n^2-3n+1\}$$

$$= \frac{12\mu_2^4(n-1)(n+3)}{n^4}.$$

Hence

$$\beta_1 = \frac{M_3^2}{M_2^3} = \frac{8}{n-1}, \quad \beta_2 = \frac{M_4}{M_2^2} = \frac{3(n+3)}{n-1)},$$

$$\therefore 2\beta_2 - 3\beta_1 - 6 = \frac{1}{n-1}\{6(n+3)-24-6(n-1)\} = 0.$$

Consequently a curve of Prof. Pearson's Type III may he expected to fit the distribution of s^2.

The equation referred to an origin at the zero end of the curve will be

$$y = Cx^p e^{-\gamma x},$$

where

$$\gamma = 2\frac{M_2}{M_3} = \frac{4\mu_2^2(n-1)n^3}{8n^2\mu_2^2(n-1)} = \frac{n}{2\mu_2}$$

and

$$p = \frac{4}{\beta_1} - 1 = \frac{n-1}{2} - 1 = \frac{n-3}{2}.$$

Consequently the equation becomes

$$y = Cx^{\frac{n-3}{2}} e^{-\frac{nx}{2\mu_2}},$$

which will give the distribution of s^2.

The area of this curve is $C\int_0^\infty x^{\frac{n-3}{2}} e^{-\frac{nx}{2\mu_2}}\,dx = I$ (say). The first moment coefficient about the end of the range will therefore be

$$\frac{C\int_0^\infty x^{\frac{n-1}{2}} e^{-\frac{nx}{2\mu_2}}\,dx}{I} = \frac{\left[C\frac{-2\mu_2}{n} x^{\frac{n-1}{2}} e^{-\frac{nx}{2\mu_2}}\right]_{x=0}^{x=\infty}}{I} + \frac{C\int_0^\infty \frac{n-1}{n}\mu_2 x^{\frac{n-3}{2}} e^{-\frac{nx}{2\mu_2}}\,dx}{I}.$$

The first part vanishes at each limit and the second is equal to

$$\frac{\frac{n-1}{n}\mu_2 I}{I} = \frac{n-1}{n}\mu_2.$$

and we see that the higher moment coefficients will he formed by multiplying successively by $\frac{n+1}{n}\mu_2$, $\frac{n+3}{n}\mu_2$ etc., just as appeared to he the law of formation of M_2', M_3', M_4', etc.

Hence it is probable that the curve found represents the theoretical distribution of s^2; so that although we have no actual proof we shall assume it to do so in what follows.

The distribution of s may he found from this, since the frequency of s is equal to that of s^2 and all that we must do is to compress the base line suitably.

Now if $\quad y_1 = \phi(s^2)$ be the frequency curve of s^2
and $\quad y_2 = \psi(s)$ be the frequency curve of s,
then

$$\begin{aligned} y_1 d(s^2) &= y_2 ds, \\ y_2 ds &= 2y_1 s ds, \\ \therefore y_2 &= 2sy_1. \end{aligned}$$

Hence

$$y_2 = 2Cs(s^2)^{\frac{n-3}{2}} e^{-\frac{ns^2}{2\mu_2}}.$$

is the distribution of s.

This reduces to

$$y_2 = 2Cs^{n-2} e^{-\frac{ns^2}{2\sigma^2}}.$$

Hence $y = Ax^{n-2}e^{-\frac{s^2}{2\mu_2}}$ will give the frequency distribution of standard deviations of samples of n, taken out of a population distributed normally with standard deviation σ^2. The constant A may he found by equating the area of the curve as follows:

$$\text{Area} = A\int_0^\infty x^{n-2} e^{-\frac{nx^2}{2\sigma^2}}\,dx. \qquad \left(\text{Let } I_p \text{ represent } \int_0^\infty x^p e^{-\frac{-nx^2}{2\sigma^2}}\,dx.\right)$$

Then

$$\begin{aligned} I_p &= \frac{\sigma^2}{n}\int_0^\infty x^{p-1}\frac{d}{dx}\left(-e^{-\frac{nx^2}{2\sigma^2}}\right)dx \\ &= \frac{\sigma^2}{n}\left[-x^{p-1}e^{-\frac{nx^2}{2\sigma^2}}\right]_{x=0}^{x=\infty} + \frac{\sigma_2}{n}(p-1)\int_0^\infty x^{p-2}e^{-\frac{x^2}{2\sigma^2}}\,dx \\ &= \frac{\sigma^2}{n}(p-1)I_{p-2}, \end{aligned}$$

since the first part vanishes at both limits.

By continuing this process we find

$$I_{n-2} = \left(\frac{\sigma^2}{n}\right)^{\frac{n-2}{2}} (n-3)(n-5)\dots 3.1 I_0$$

or

$$I_{n-2} = \left(\frac{\sigma^2}{n}\right)^{\frac{n-2}{2}} (n-3)(n-5)\dots 4.2 I_1$$

according n is even or odd.

But I_0 is

$$\int_0^\infty e^{-\frac{nx^2}{2\sigma^2}} dx = \sqrt{\left(\frac{\pi}{2n}\right)}\sigma,$$

and I_1 is

$$\int_0^\infty xe^{-\frac{nx^2}{2sigma^2}} dx = \left[-\frac{\sigma^2}{n}e^{-\frac{nx^2}{2\sigma^2}}\right]_{x=0}^{x=\infty} = \frac{\sigma^2}{n}.$$

Hence if n be even,

$$A = \frac{\text{Area}}{(n-3)(n-5)\dots 3.1\sqrt{\left(\frac{\pi}{2}\right)}\left(\frac{\sigma^2}{n}\right)^{\frac{n-1}{2}}},$$

while is n be odd

$$A = \frac{\text{Area}}{(n-3)(n-5)\dots 4.2\left(\frac{\sigma^2}{n}\right)^{\frac{n-1}{2}}}.$$

Hence the equation may be written

$$y = \frac{N}{(n-3)(n-5)\dots 3.1}\sqrt{\left(\frac{2}{\pi}\right)}\left(\frac{n}{\sigma^2}\right)^{\frac{n-1}{2}} x^{n-2}e^{-\frac{nx^2}{2\sigma^2}} \ (n \text{ even})$$

or

$$y = \frac{N}{(n-3)(n-5)\dots 4.2}\left(\frac{n}{\sigma^2}\right)^{\frac{n-1}{2}} x^{n-2}e^{-\frac{nx^2}{2\sigma^2}} \ (n \text{ odd})$$

where N as usual represents the total frequency.

Section II

To show that there is no correlation between (a) the distance of the mean of a sample from the mean of the population and (b) the standard deviation of a sample with normal distribution.

(1) Clearly positive and negative positions of the mean of the sample are equally likely, and hence there cannot be correlation between the absolute value of the distance of the mean from the mean of the population and the standard

deviation, but (2) there might be correlation between the square of the distance and the square of the standard deviation. Let

$$u^2 = \left(\frac{S(x_1)}{n}\right)^2 \quad \text{and} \quad s^2 = \frac{S(x_1^2)}{n} - \left(\frac{S(x_1)}{n}\right)^2.$$

Then if m_1', M_1' be the mean values of u^2 and s^z, we have by the preceding part

$$M_1' = \mu_2 \frac{(n-1)}{n} \quad \text{and} \quad m_1' = \frac{\mu_2}{n}.$$

Now

$$\begin{aligned}
u^2 s^2 &= \frac{S(x_1^2)}{n}\left(\frac{S(x_1)}{n}\right)^2 - \left(\frac{S(x_1)}{n}\right)^4 \\
&= \left(\frac{S(x_1^2)}{n}\right)^2 + 2\frac{S(x_1x_2).S(x_1^2)}{n^3} - \frac{S(x_1^4)}{n^4} - \frac{6S(x_1^2x_2^2)}{n_4} \\
&\qquad - \text{other terms of odd order which will vanish on summation.}
\end{aligned}$$

Summing for all values and dividing by the number of cases we get

$$R_{u^2 s_2}\sigma_{u^2}\sigma_{s^2} + m_1 M_1 = \frac{\mu_4}{n_2} + \mu_2^2\frac{(n-1)}{n^2} - \frac{\mu_4}{n_3} - 3\mu_2^2\frac{(n-1)}{n^3},$$

where $R_{u^2s^2}$ is the correlation between u^2 and s^2.

$$R_{u^2 s_2}\sigma_{u^2}\sigma_{s^2} + \mu_2^2\frac{(n-1)}{n^2} = \mu_2^2\frac{(n-1)}{n^3}\{3+n-3\} = \mu_2^2\frac{(n-1)}{n^2}.$$

Hence $R_{u^2 s_2}\sigma_{u^2}\sigma_{s^2} = 0$, or there is no correlation between u^2 and s^2.

Section III

To find the equation representing the frequency distribution of the means of samples of n drawn from a normal population, the mean being expressed in terms of the standard deviation of the sample.

We have $y = \frac{C}{\sigma^{n-1}} s^{n-2} e^{-\frac{nx^2}{2\sigma^2}}$ as the equation representing the distribution of s, the standard deviation of a sample of n, when the samples are drawn from a normal population with standard deviation s.

Now the means of these samples of n are distributed according to the equation[1]

$$y = \frac{\sqrt{(n)}N}{\sqrt{(2\pi)}\sigma} e^{-\frac{nx^2}{2\sigma^2}},$$

and we have shown that there is no correlation between x, the distance of the mean of the sample, and s, the standard deviation of the sample.

[1] Airy, *Theory of Errors of Observations*, Part II, §6.

Now let us suppose x measured in terms of s, i.e. let us find the distribution of $z = x/s$.

If we have $y_1 = \phi(x)$ and $y_2 = \psi(z)$ as the equations representing the frequency of x and of z respectively, then

$$y_1 dx = y_2 dz = y_3 \frac{dx}{s},$$
$$\therefore y_2 = s y_1.$$

Hence

$$y = \frac{N\sqrt{(n)}s}{\sqrt{(2\pi)}\sigma} e^{-\frac{ns^2z^2}{2\sigma^2}}$$

is the equation representing the distribution of z for samples of n with standard deviation s.

Now the chance that s lies between s and $s + ds$ is

$$\frac{\int_s^{s+ds} \frac{C}{\sigma^{n-1}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}{\int_0^\infty \frac{C}{\sigma^{n-1}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}$$

which represents the N in the above equation.

Hence the distribution of z due to values of s which lie between s and $s+ds$ is

$$y = \frac{\int_s^{s+ds} \frac{C}{\sigma^n} \sqrt{\left(\frac{n}{2\pi}\right)} s^{n-1} e^{-\frac{ns^2(1+z^2)}{2\sigma^2}} ds}{\int_0^\infty \frac{C}{\sigma^{n-1}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds} = \sqrt{\left(\frac{n}{2\pi}\right)} \frac{\int_s^{s+ds} \frac{C}{\sigma^n} s^{n-1} e^{-\frac{ns^2(1+z^2)}{2\sigma^2}} ds}{\int_0^\infty \frac{C}{\sigma^{n-2}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}$$

and summing for all values of s we have as an equation giving the distribution of z

$$y = \frac{\sqrt{\left(\frac{n}{2\pi}\right)}}{\sigma} \frac{\int_s^{s+ds} \frac{C}{\sigma^n} s^{n-1} e^{-\frac{ns^2(1+z^2)}{2\sigma^2}} ds}{\int_0^\infty \frac{C}{\sigma^{n-2}} s^{n-2} e^{-\frac{ns^2}{2\sigma^2}} ds}.$$

By what we have already proved this reduces to

$$y = \frac{1}{2} \frac{n-2}{n-3} . \frac{n-4}{n-5} \dots \frac{5}{4} . \frac{3}{2} (1+z^2)^{-\frac{1}{2}n}, \quad \text{if } n \text{ be odd}$$

and to

$$y = \frac{1}{2} \frac{n-2}{n-3} . \frac{n-4}{n-5} \dots \frac{4}{3} . \frac{2}{21} (1+z^2)^{-\frac{1}{2}n}, \quad \text{if } n \text{ be even}$$

Since this equation is independent of σ it will give the distribution of the distance of the mean of a sample from the mean of the population expressed in terms of the standard deviation of the sample for any normal population.

Section IV. Some Properties of the Standard Deviation Frequency Curve

By a similar method to that adopted for finding the constant we may find the mean and moments: thus the mean is at I_{n-1}/I_{n-2},
which is equal to

$$\frac{n-2}{n-3}.\frac{n-4}{n-5}\cdots\frac{2}{1}\sqrt{\left(\frac{2}{\pi}\right)}\frac{\sigma}{\sqrt{n}}, \quad \text{if } n \text{ be even,}$$

or

$$\frac{n-2}{n-3}.\frac{n-4}{n-5}\cdots\frac{3}{2}\sqrt{\left(\frac{\pi}{2}\right)}\frac{\sigma}{\sqrt{n}}, \quad \text{if } n \text{ be odd .}$$

The second moment about the end of the range is

$$\frac{I_n}{I_{n-2}} = \frac{(n-1)\sigma^2}{n}.$$

The third moment about the end of the range is equal to

$$\begin{aligned}\frac{I_{n+1}}{I_{n-2}} &= \frac{I_{n+1}}{I_{n-1}}.\frac{In-1}{I_{n-2}} \\ &= \sigma^2 \times \text{the mean.}\end{aligned}$$

The fourth moment about the end of the range is equal to

$$\frac{I_{n+2}}{I_{n-2}} = \frac{(n-1)(n+1)}{n^2}\sigma^4.$$

If we write the distance of the mean from the end of the range $D\sigma/\sqrt{n}$ and the moments about the end of the range ν_1, ν_2, etc.,
then

$$\nu_1 = \frac{D\sigma}{\sqrt{n}}, \quad \nu_2 = \frac{n-1}{n}\sigma_2, \quad \nu_3 = \frac{D\sigma^3}{\sqrt{n}}, \quad \nu_4 = \frac{N^2-1}{n}\sigma^4.$$

From this we get the moments about the mean:

$$\begin{aligned}\mu_2 &= \frac{\sigma^2}{n}(n-1-D^2), \\ \mu_3 &= \frac{\sigma^3}{n\sqrt{n}}\{nD-3(n-1)D+2D^2\} = \frac{\sigma^3 D}{n\sqrt{n}}\{2D^2-2n+3\}, \\ \mu_4 &= \frac{\sigma^2}{n^2}\{n^2-1-4D^2n+6(n-1)D^2-3D^4\} \\ &= \frac{\sigma^4}{n^2}\{n^2-1-D^2(3D^2-2n+6)\}.\end{aligned}$$

It is of interest to find out what these become when n is large.

In order to do this we must find out what is the value of D.

Now Wallis's expression for π derived from the infinite product value of $\sin x$ is

$$\frac{\pi}{2}(2n+1) = \frac{2^2.4^2.6^2\ldots(2n)^2}{1^2 3^2 5^2 \ldots (2n-1)^2}.$$

If we assume a quantity $\theta \left(= a_0 + \frac{a_1}{n} + \text{etc.}\right)$ which we may add to the $2n+1$ in order to make the expression approximate more rapidly to the truth, it is easy to show that $\theta = -\frac{1}{2} + \frac{1}{16n} - \text{etc.}$, and we get[2]

$$\frac{\pi}{2}\left(2n + \frac{1}{2} + \frac{1}{16n}\right) = \frac{2^2.4^2.6^2\ldots(2n)^2}{1^2 3^2 5^2 \ldots (2n-1)^2}.$$

From this we find that whether n be even or odd D^2 approximates to $n - \frac{3}{2} + \frac{1}{8n}$ when n is large.

Substituting this value of D we get

$$\mu_2 = \frac{\sigma^2}{2n}\left(1 - \frac{1}{4n}\right),\ \mu_2 = \frac{\sigma^3\sqrt{\left(1 - \frac{3}{2n} + \frac{1}{16n^2}\right)}}{4n^2},\ \mu_4 = \frac{3\sigma_4}{4n^2}\left(1 + \frac{1}{2n} - \frac{1}{16n^2}\right).$$

Consequently the value of the standard deviation of a standard deviation which we have found $\left(\frac{\sigma}{\sqrt{(2n)}\sqrt{\{1-(1/4n)\}}}\right)$ becomes the same as that found for the normal curve by Prof. Pearson $\{\sigma/(2n)\}$ when n is large enough to neglect the $1/4n$ in comparison with 1.

Neglecting terms of lower order than $1/n$, we find

$$\beta_1 = \frac{2n-3}{n(4n-3)},\quad \beta)_2 = 3\left(1 - \frac{1}{2n}\right)\left(1 + \frac{1}{2n}\right).$$

Consequently, as n increases, β_2 very soon approaches the value 3 of the normal curve, but β_1 vanishes more slowly, so that the curve remains slightly skew.

Diagram I shows the theoretical distribution of the standard deviations found from samples of 10.

Section V. Some Properties of the Curve

$$y = \frac{n-2}{n-3}.\frac{n-4}{n-5}\cdots\left(\begin{array}{l}\frac{4}{3}.\frac{2}{\pi}\text{ if } n \text{ be even}\\ \frac{5}{4}.\frac{3}{2}.\frac{1}{2}\text{ if } n \text{ be odd}\end{array}\right)(1+z^2)^{-\frac{1}{2}n}$$

Writing $z = \tan\theta$ the equation becomes $y = \frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \cos^n\theta$, which affords an easy way of drawing the curve. Also $dz = d\theta/\cos^2\theta$.

[2]This expression will be found to give a much closer approximation to π than Wallis's

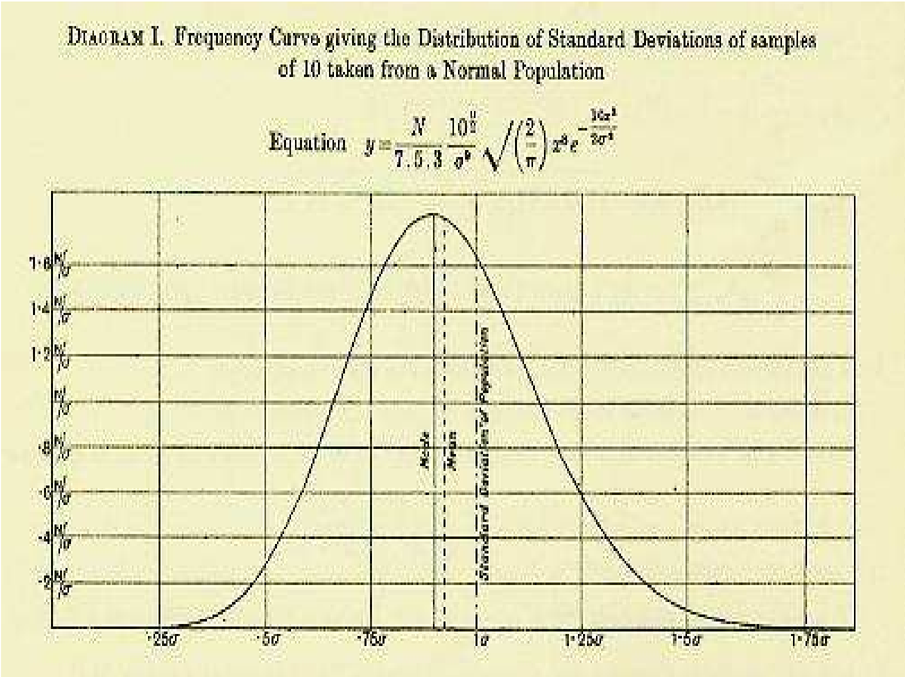


DIAGRAM I. Frequency Curve giving the Distribution of Standard Deviations of samples of 10 taken from a Normal Population

Hence to find the area of the curve between any limits we must find

$$
\begin{aligned}
&\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.}\times\int\cos^{n-2}\theta d\theta \\
=&\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.}\left\{\frac{n-3}{n-2}\int\cos^{n-4}\theta d\theta+\left[\frac{cos^{n-3}\theta\sin\theta}{n-2}\right]\right\} \\
=&\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.}\int\cos^{n-4}\theta d\theta+\frac{1}{n-3}\frac{n-4}{n-5}\ldots\text{etc.}[\cos^{n-3}\theta\sin\theta],
\end{aligned}
$$

and by continuing the process the integral may he evaluated.

For example, if we wish to find the area between 0 and θ for $n = 8$ we have

$$
\begin{aligned}
\text{Area} &= \frac{6}{5}.\frac{4}{3}.\frac{2}{1}.\frac{1}{\pi}\int_0^\theta\cos^6\theta d\theta \\
&= \frac{4}{3}.\frac{2}{\pi}\int_0^\theta\cos^4\theta d\theta+\frac{1}{5}.\frac{4}{3}.\frac{2}{\pi}\cos^5\theta\sin\theta \\
&= \frac{\theta}{\pi}+\frac{1}{\pi}\cos\theta\sin\theta+\frac{1}{3}.\frac{2}{\pi}\cos^3\theta\sin\theta+\frac{1}{5}.\frac{4}{3}.\frac{2}{\pi}\cos^5\theta\sin\theta
\end{aligned}
$$

and it will be noticed that for $n = 10$ we shall merely have to add to this same expression the term $\frac{1}{7}.\frac{6}{5}.\frac{4}{3}.\frac{2}{\pi}\cos^7\theta\sin\theta$.

The tables at the end of the paper give the area between $-\infty$ and z

$$\left(\text{or } \theta = -\frac{\pi}{2} \text{ and } \theta = \tan^{-1} z\right).$$

This is the same as $0.5 +$ the area between $\theta = 0$, and $\theta = \tan^{-1} z$, and as the whole area of the curve is equal to 1, the tables give the probability that the mean of the sample does not differ by more than z times the standard deviation of the sample from the mean of the population.

The whole area of the curve is equal to

$$\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} \cos^{n-2}\theta d\theta$$

and since all the parts between the limits vanish at both limits this reduces to 1.

Similarly, the second moment coefficient is equal to

$$\begin{aligned}
&\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} \cos^{n-2}\theta \tan^2\theta d\theta \\
&\quad = \frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} (\cos^{n-4}\theta - \cos^{n-2}\theta) d\theta \\
&\quad = \frac{n-2}{n-3} - 1 = \frac{1}{n-3}.
\end{aligned}$$

Hence the standard deviation of the curve is $1/\sqrt{(n-3)}$. The fourth moment coefficient is equal to

$$\begin{aligned}
&\frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} \cos^{n-2}\theta \tan^4\theta d\theta \\
&\quad = \frac{n-2}{n-3}.\frac{n-4}{n-5}\ldots\text{etc.} \times \int_{-\frac{1}{2}\pi}^{+\frac{1}{2}\pi} (\cos^{n-6}\theta - 2\cos^{n-4}\theta + \cos^{n-2}\theta) d\theta \\
&\quad = \frac{n-2}{n-3}.\frac{n-4}{n-5} - \frac{2(n-2)}{n-3} + 1 = \frac{3}{(n-3)(n-5)}.
\end{aligned}$$

The odd moments are of course zero, an the curve is symmetrical, so

$$\beta_1 = 0, \quad \beta_2 = \frac{3(n-3)}{n-5} = 3 + \frac{6}{n-5}.$$

Hence as it increases the curve approaches the normal curve whose standard deviation is $1/\sqrt{(n-3)}$.

β_2, however, is always greater than 3, indicating that large deviations are mere common than in the normal curve.

I have tabled the area for the normal curve with standard deviation $1/\sqrt{7}$ so as to compare, with my curve for $n = 10$[3]. It will be seen that odds laid

[3]See p. 29

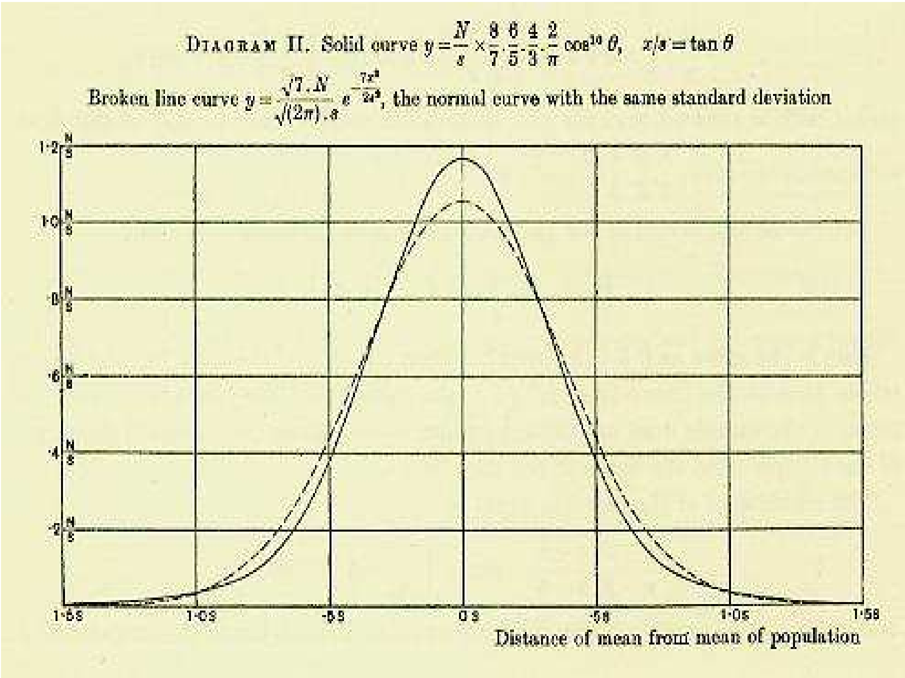


DIAGRAM II. Solid curve $y=\frac{N}{s}\times\frac{8}{7}\cdot\frac{6}{5}\cdot\frac{4}{3}\cdot\frac{2}{\pi}\cos^{10}\theta$, $x/s=\tan\theta$

Broken line curve $y=\frac{\sqrt{7}.N}{\sqrt{(2\pi)}.s}e^{-\frac{7x^2}{2s^2}}$, the normal curve with the same standard deviation

according to either table would not seriously differ till we reach $z=0.8$, where the odds are about 50 to 1 that the mean is within that limit: beyond that the normal curve gives a false feeling of security, for example, according to the normal curve it is 99,986 to 14 (say 7000 to 1) that the mean of the population lies between $-\infty$ and $+1.3s$, whereas the real odds are only 99,819 to 181 (about 550 to 1).

Now 50 to 1 corresponds to three times the probable error in the normal curve and for most purposes it 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 he used as foundations for finding values for larger samples.[4]

The table for $n=2$ can be readily constructed by looking out $\theta=\tan^{-1}z$ in Chambers's tables and then $0.5+\theta/\pi$ gives the corresponding value.

Similarly $\frac{1}{2}\sin\theta+0.5$ 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 $-z$ 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

[4]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.

moment coefficient is

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

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

Section VI. Practical Test of the foregoing Equations

Before I bad 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. Macdonnell (*Biometrika*, 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[5] 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 $\beta_1 = 0.0026$, $\beta_2 = 3.176$, and for left middle finger lengths $\beta_1 = 0.0030$, $\beta_2 = 3.140$, and in consequence there is a tendency for a certain number of larger standard deviations to occur than if the distributions wore normal. This, however, appears to make very little difference to the distribution of z.

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 wore originally grouped in millimetres, but unfortunately I did not at the time see the importance of having a smaller unit and condensed them into 2 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 0.25. (2) The standard deviations occur as the square roots of the following types of numbers: n, $n+0.10$, $n+0.25$, $n+0.50$, $n+0.69$, $2n+0.75$.

(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

[5] I hope to publish the results of the correlation work shortly.

only occur if the mean differs by a whole number from the group we take as origin, while $\sqrt{1.69}$ will only occur when the mean is at $n \pm 0.25$.

(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 he 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 0.436, when 3 fall in one group and 1 in an adjacent group, or 0.50 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 he 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 neutralize the 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 he considered satisfactory.

The following are the figures for height:

Mean value of standard deviations:	Calculated	2.027 ± 0.02
	Observed	2.026
	Difference =	-0.001
Standard deviation of standard deviations:	Calculated	0.8558 ± 0.015
	Observed	0.9066
	Difference	$+0.0510$

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

Scale in terms of standard deviations of population																		
Calculated frequency	$1\frac{1}{2}$	$10\frac{1}{2}$	27	$45\frac{1}{2}$	$64\frac{1}{2}$	$78\frac{1}{2}$	87	88	$81\frac{1}{2}$	71	58	45	33	23	15	$9\frac{1}{2}$	$5\frac{1}{2}$	7
Observed frequency	3	$14\frac{1}{2}$	$24\frac{1}{2}$	$37\frac{1}{2}$	107	67	73	77	$77\frac{1}{2}$	64	$\frac{1}{2}$	$49\frac{1}{2}$	35	28	$12\frac{1}{2}$	9	$11\frac{1}{2}$	7
Difference	$+1\frac{1}{2}$	$+4$	$-2\frac{1}{2}$	-8	$+42\frac{1}{2}$	$-11\frac{1}{2}$	-14	-11	-4	-7	$-5\frac{1}{2}$	$+4\frac{1}{2}$	$+2$	$+5$	$-2\frac{1}{2}$	$-\frac{1}{2}$	$+6$	0

Whence $\chi^2 = 48.06$, $P = 0.00006$ (about).

In tabling the observed frequency, values between 0.0125 and 0.0875 were included in one group, while between 0.0875 and 0.012.5 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 0.5117 in terms of the standard deviation of the population, and they wore therefore divided over the groups 0.4 to 0.5 and 0.5 to 0.6. Had they all been counted in groups 0.5 to 0.6 χ^2 would have fallen to 20.85 and P would have risen to 0.03. 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 0.1; 0.04, 0.05 and 0.06, being divided between the two groups on either side.

The calculated value for the standard deviation of the frequency curve was 1 (± 0.0171), while the observed was 1.030. 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 $1/\sqrt{1500}$ 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: $y = \frac{2N}{\pi}\cos^4\theta$, $z = \tan\theta$

Scale of z															
Calculated frequency	5	$9\frac{1}{2}$	$13\frac{1}{2}$	$34\frac{1}{2}$	$44\frac{1}{2}$	$78\frac{1}{2}$	119	141	$78\frac{1}{2}$	$44\frac{1}{2}$	$34\frac{1}{2}$	$13\frac{1}{2}$	$13\frac{1}{2}$	$9\frac{1}{2}$	5
Observed frequency	9	$14\frac{1}{2}$	$11\frac{1}{2}$	33	$43\frac{1}{2}$	$70\frac{1}{2}$	$119\frac{1}{2}$	$151\frac{1}{2}$	122	$67\frac{1}{2}$	49	$26\frac{1}{2}$	16	10	6
Difference	$+4$	$+4$	-2	-2	$-1\frac{1}{2}$	-1	-8	$+\frac{1}{2}$	$+10\frac{1}{2}$	$+3$	-11	$+4\frac{1}{2}$	-8	$+2\frac{1}{2}$	$+\frac{1}{2}$

Whence $\chi^2 = 12.44$, $P = 0.56$.

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])c expected; while this curve is exposed to the full errors of random sampling, its constants having been calculated quite apart from the observations.

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[6] and IV give the standard deviations of the z's for the set of samples. The results are as follows:

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

Mean value of standard deviations:	Calculated	2.186 ± 0.023
	Observed	2.179
	Difference =	-0.007
Standard deviation of standard deviations:	Calculated	0.9224 ± 0.016
	Observed	0.9802
	Difference =	$+0.0578$

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

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

Whence $\chi^2 = 21.80$, $P = 0.19$.

Value of standard deviation:	Calculated	$1(\pm 0.017)$
	Observed	0.982
	Difference =	-0.018

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

Scale of z															
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 $\chi^2 = 7.39$, $P = 0.92$.

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 although both will have greater variability, yet they will tend to counteract one another, a mean deriving largely from the general mean tending to be divided by a larger standard deviation. Consequently, I believe that the table given in Section VII below 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

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

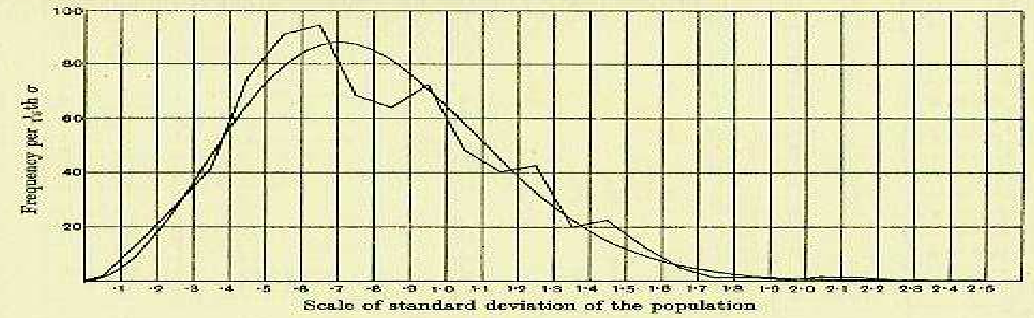
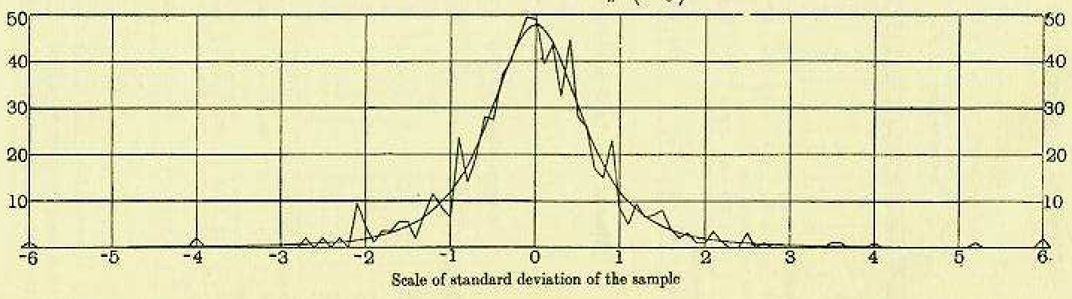


DIAGRAM IV. Comparison of the Theoretical Frequency Curve $y = \frac{1500}{\pi}\left(1+\frac{x^2}{s^2}\right)^{-2}$, with an actual Sample of 750 Cases



Section VII. Tables of

$$\frac{n-2}{n-3}\cdot\frac{n-4}{n-5}\cdots\left(\begin{array}{l}\frac{3}{2}\cdot\frac{1}{2}\,n\ \textbf{odd}\\ \frac{2}{1}\cdot\frac{1}{\pi}\,n\ \textbf{even}\end{array}\right)\int_{-\frac{1}{2}\pi}^{\tan^{-1}z}\cos^{n-2}\theta d\theta$$

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 $\left(\frac{\sqrt{7}}{\sqrt{(2\pi)}}\int_{-\infty}^{x} e^{-\frac{7x^2}{2}}dx\right)$
0.1	0.5633	0.5745	0.5841	0.5928	0.6006	0.60787	0.61462	0.60411
0.2	0.6241	0.6458	0.6634	0.6798	0.6936	0.70705	0.71846	0.70159
0.3	0.6804	0.7096	0.7340	0.7549	0.7733	0.78961	0.80423	0.78641
0.4	0.7309	0.7657	0.7939	0.8175	0.8376	0.85465	0.86970	0.85520
0.5	0.7749	0.8131	0.8428	0.8667	0.8863	0.90251	0.91609	0.90691
0.6	0.8125	0.8518	0.8813	0.9040	0.9218	0.93600	0.94732	0.94375
0.7	0.8440	0.8830	0.9109	0.9314	0.9468	0.95851	0.96747	0.96799
0.8	0.8701	0.9076	0.9332	0.9512	0.9640	0.97328	0.98007	0.98253
0.9	0.8915	0.9269	0.9498	0.9652	0.9756	0.98279	0.98780	0.99137
1.0	0.9092	0.9419	0.9622	0.9751	0.9834	0.98890	0.99252	0.99820
1.1	0.9236	0.9537	0.9714	0.9821	0.9887	0.99280	0.99539	0.99926
1.2	0.9354	0.9628	0.9782	0.9870	0.9922	0.99528	0.99713	0.99971
1.3	0.9451	0.9700	0.9832	0.9905	0.9946	0.99688	0.99819	0.99986
1.4	0.9451	0.9756	0.9870	0.9930	0.9962	0.99791	0.99885	0.99989
1.5	0.9598	0.9800	0.9899	0.9948	0.9973	0.99859	0.99926	0.99999
1.6	0.9653	0.9836	0.9920	0.9961	0.9981	0.99903	0.99951	
1.7	0.9699	0.9864	0.9937	0.9970	0.9986	0.99933	0.99968	
1.8	0.9737	0.9886	0.9950	0.9977	0.9990	0.99953	0.99978	
1.9	0.9970	0.9904	0.9959	0.9983	0.9992	0.99967	0.99985	
2.0	0.9797	0.9919	0.9967	0.9986	0.9994	0.99976	0.99990	
2.1	0.9821	0.9931	0.9973	0.9989	0.9996	0.99983	0.99993	
2.2	0.9841	0.9941	0.9978	0.9992	0.9997	0.99987	0.99995	
2.3	0.9858	0.9950	0.9982	0.9993	0.9998	0.99991	0.99996	
2.4	0.9873	0.9957	0.9985	0.9995	0.9998	0.99993	0.99997	
2.5	0.9886	0.9963	0.9987	0.9996	0.9998	0.99995	0.99998	
2.6	0.9898	0.9967	0.9989	0.9996	0.9999	0.99996	0.99999	
2.7	0.9908	0.9972	0.9989	0.9997	0.9999	0.99997	0.99999	
2.8	0.9916	0.9975	0.9989	0.9998	0.9999	0.99998	0.99999	
2.9	0.9924	0.9978	0.9989	0.9998	0.9999	0.99998	0.99999	
3.0	0.9931	0.9981	0.9989	0.9998	—	0.99999	—	

Explanation of Tables

The tables give the probability that the value of the mean, measured from the mean of the population, in terms of the standard deviation of the sample, will lie between $-\infty$ and z. Thus, to take the table for samples of 6, the probability of the mean of the population lying between $-\infty$ and once the standard deviation of the sample is 0.9622, the odds are about 24 to 1 that the mean of the population lies between these limits.

The probability is therefore 0.0378 that it is greater than once the standard deviation and 0.07511 that it lies outside ±1.0 times the standard deviation.

Illustration of Method

Illustration I. As an instance of the kind of use which may be made of the tables, I take the following figures from a table by A. R. Cushny and A. R. Peebles in the *Journal of Physiology* for 1904, showing the different effects of the optical isomers of hyoscyamine hydrobromide in producing sleep. The average number of hours' sleep gained by the use of the drug is tabulated below.

The conclusion arrived at was that in the usual doses 2 was, but 1 was not, of value as a soporific.

Additional hours' sleep gained by the use of hyoscyamine hydrobromide

Patient		1 (Dextro-)		2 (Laevo-)		Difference (2 − 1)
1		+0.7		+1.9		+1.2
2		−1.6		+0.8		+2.4
3		−0.2		+1.1		+1.3
4		−1.2		+0.1		+1.3
5		−0.1		−0.1		0
6		+3.4		+4.4		+1.0
7		+3.7		+5.5		+1.8
8		+0.8		+1.6		+0.8
9		0		+4.6		+4.6
10		+2.0		+3.4		+1.4
	Mean	+0.75	Mean	+2.33	Mean	+1.58
	S.D.	1.70	S.D.	1.90	S.D.	1.17

First let us see what is the probability that 1 will on the average give increase of sleep; i.e. what is the chance that the mean of the population of which these experiments are a sample is positive. +0.75/1.70 = 0.44, and looking out $z =$ 0.44 in the table for ten experiments we find by interpolating between 0.8697 and 0.9161 that 0.44 corresponds to 0.8873, or the odds are 0.887 to 0.113 that the mean is positive.

That is about 8 to 1, and would correspond to the normal curve to about 1.8 times the probable error. It is then very likely that 1 gives an increase of sleep, but would occasion no surprise if the results were reversed by further experiments.

If now we consider the chance that 2 is actually a soporific we have the mean inclrease of sleep = 2.33/1.90 or 1.23 times the S.D. From the table the probability corresponding to this is 0.9974, i.e. the odds are nearly 400 to 1 that such is the case. This corresponds to about 4.15 times the probable error in the normal curve. But I take it that the real point of the authors was that 2 is better than 1. This we must t4est by making a new series, subtracting 1 from 2. The mean values of this series is +1.38, while the S.D. is 1.17, the mean value being +1.35 times the S.D. From the table, the probability is 0.9985, or the odds are about 666 to one that 2 is the better soporific. The low value of the s.d. is probably due to the different drugs reacting similarly on the same patient, so that there is correlation between the results.

Of course odds of this kind make it almost certain that 2 is the better soporific, and in practical life such a high probability is in most matters considered as a certainty.

Illustration II. Cases where the tables will be useful are not uncommon in agricultural work, and they would be more numerous if the advantages of being able to apply statistical reasoning were borne in mind when planning the experiments. I take the following instances from the accounts of the Woburn farming experiments published yearly by Dr Voelcker in the *Journal of the Agricultural Soceity*.

A short series of pot culture experiments were conducted in order to determine the casues which lead to the production of Hard (glutinous) wheat or Soft (starchy) wheat. In three successive years a bulk of seed corn of one variety was picked over by hand and two samples were selected, one consisting of "hard" grains avid the other of "soft". Some of each of them were planted in both heavy and light soil and the resulting crops wore weighed and examined for hard and soft corn.

The conclusion drawn was that the effect of selecting the seed was negligible compared with the influence of the soil.

This conclusion was thoroughly justified, theheavy soul producing in each case nearly 100% of hard corn, but still the effect of selecting the seed could just be traced in each year.

But a curious point, to which Dr Voelcker draws attention in the second year's report, is that the soft seeds produced the higher yield of both corn and straw. In view of the well-known fact that the *varieties* which have a high yield tend to produce soft corn, it is interesting to see how much evidence the experiments afford as to the correlation between softness and fertility in the same *variety*.

Further, Mr Hooker[7] has shown that the yield of wheat in one year is largely determined by the weather during the preceding year. Dr Voelcker's results may afford a clue as to the way in which the seed id affected, and would almost justify the selection of particillar soils for growing wheat.[8]

Th figures are as follows, the yields being expressed in grammes per pot:

Year	1899		1900		1901			Standard	
Soil	Light	Heavy	Light	Heavy	Light	Heavy	Average	deviation	z
Yield of corn from soft seed	7.55	8.89	14.81	13.55	7.49	15.39	11.328		
Yield of corn from hard seed	7.27	8.32	13.81	13.36	7.97	13.13	10.643		
Difference	+0.58	+0.57	+1.00	+0.19	−0.49	+2.26	+0.685	0.778	0.88
Yield of straw from soft seed	12.81	12.87	22.22	20.21	13.97	22.57	17.442		
Yield of straw from hard seed	10.71	12.48	21.64	20.26	11.71	18.96	15.927		
Difference	+2.10	+0.39	+0.78	−0.05	+2.66	+3.61	+1.515	1.261	1.20

If we wish to laid the odds that the soft seed will give a better yield of corn on the average, we divide, the average difference by the standard deviation,

[7] *Journal of the Royal Statistical Society*, 1897

[8] And perhaps a few experiments to see whether there is a correlation between yield and "mellowness" in barley.

giving us

$$z = 0.88.$$

Looking this up in the table for $n = 6$ we find $p = 0.9465$ or the odds are 0.9465 to 0.0535 about 18 to 1.

Similarly for straw $z = 1.20$, $p = 0.9782$, and the odds are about 45 to 1.

In order to see whether such odds are sufficient for a practical man to draw a definite conclusion, I take another act of experiments in which Dr Voelcker compares the effects of different artificial manures used with potatoes on a large scale.

The figures represent the difference between the crops grown with the rise of sulphate of potash and kailit respectively in both 1904 and 1905:

$$\left.\begin{array}{lrrrclrrr} & \text{cwt.} & \text{qr.} & \text{lb.} & & \text{ton} & \text{cwt.} & \text{qr.} & \text{lb.} \\ 1904+ & 10 & 3 & 20 & :+ & 1 & 10 & 1 & 26 \\ 1905+ & 6 & 0 & 3 & :+ & & 13 & 2 & 8 \end{array}\right\}\ \text{(two experiments in each year)}$$

The average gain by the use of sulphate of potash was 15.25 cwt. and the S.D. 9 cwt., whence, if we want the odds that the conclusion given below is right, $z = 1.7$, corresponding, when $n = 4$,to $p = 0.9698$ or odds of 32 to 1; this is midway between the odds in the former example. Dr Voelcker says: "It may now fairly be concluded that for the potato crop on light land 1 cwt. per acre of sulphate of potash is a better dressing than kailit."

Am an example of how the table should be used with caution, I take the following pot culture experiments to test whether it made any difference whether large or small seeds were sown.

Illustration III. In 1899 and in 1903 "head corn" and "tail corn" were taken from the same bulks of barley and sown in pots. The yields in grammes were as follows:

	1899	1903
Large seed ...	13.9	7.3
Small seed ...	14.4	1.4
	+0.5	+1.4

The average gain is thus 0.95 and the S.D. 0.45, giving $z = 2.1$. Now the table for $n = 2$ is not given, but if we look up the angle whose tangent is 2.1 in Chambers's tables,

$$p = \frac{\tan^{-1} 2.1}{180^\circ} + 0.5 = \frac{64^\circ 39'}{180^\circ} = 0.859,$$

so that the odds are about 6 to 1 that small corn gives a better yield than large. These odds[9] are those which would be laid, and laid rigidly, by a man whose only knowledge of the matter was contained in the two experiments. Anyone conversant with pot culture would however know that the difference between the

[9] [Through a numerical slip, now corrected, Student had given the odds as 33 to 1 and it is to this figure that the remarks in this paragraph relate.

two results would generally be greater and would correspondingly moderate the certainty of his conclusion. In point of fact a large-scale experiment confirmed this result, the small corn yielding shout 15% more than the large.

I will conclude with an example which comes beyond the range of the tables, there being eleven experiments.

To test whether it is of advantage to kiln-dry barley seed before sowing, seven varieties of barley wore sown (both kiln-dried and not kiln-dried in 1899 and four in 1900; the results are given in the table.

	Lb. head corn per acre			Price of head corn in shillings per quarter			Cwt. straw per acre			Value of crop per acre in shillings		
	N.K.D.	N.D.	Diff.	N.K.D.	N.D.	Diff.	N.K.D.	N.D.	Diff.	N.K.D.	N.D.	Diff.
	1903	2009	+106	$26\frac{1}{2}$	$\frac{1}{2}$	0	$19\frac{1}{2}$	25	$+5\frac{1}{2}$	$140\frac{1}{2}$	152	$+11\frac{1}{2}$
	1935	1915	− 20	28	$26\frac{1}{2}$	$-1\frac{1}{2}$	$22\frac{1}{2}$	24	$+1\frac{1}{2}$	$152\frac{1}{2}$	145	$-7\frac{1}{2}$
	1910	2011	+101	$29\frac{1}{2}$	$28\frac{1}{2}$	−1	23	24	+1	$158\frac{1}{2}$	161	$+2\frac{1}{2}$
1899	2496	2463	− 33	30	29	−1	23	28	+5	$204\frac{1}{2}$	$199\frac{1}{2}$	−5
	2108	2180	+ 72	$27\frac{1}{2}$	27	$-\frac{1}{2}$	$22\frac{1}{2}$	$22\frac{1}{2}$	0	162	142	+2
	1961	1925	−36	26	26	0	$19\frac{1}{2}$	$419\frac{1}{2}$	$-\frac{1}{2}$	142	$139\frac{1}{2}$	$-2\frac{1}{2}$
	2060	2122	+ 62	29	26	−3	$24\frac{1}{2}$	$22\frac{1}{2}$	$-2\frac{1}{2}$	168	155	−13
	1444	1482	+ 38	$29\frac{1}{2}$	$28\frac{1}{2}$	−1	$15\frac{1}{2}$	16	$+\frac{1}{2}$	118	‘$117\frac{1}{2}$	$-\frac{1}{2}$
1900	1612	1443	− 70	$28\frac{1}{2}$	28	$-\frac{1}{2}$	18	$17\frac{1}{2}$	$-\frac{1}{2}$	$128\frac{1}{2}$	121	$-7\frac{1}{2}$
	1316	1443	+127	30	29	−1	$14\frac{1}{2}$	$15\frac{1}{2}$	$+1\frac{1}{2}$	$109\frac{1}{2}$	$116\frac{1}{2}$	+7
	1511	1535	+ 24	$28\frac{1}{2}$	28	$-\frac{1}{2}$	17	$17\frac{1}{2}$	$+\frac{1}{2}$	120	$120\frac{1}{2}$	$+\frac{1}{2}$
Average	1841.5	1875.2	+33.7	28.45	27.55	−0.91	19.95	21.05	+1.10	145.82	144.68	+1.14
Standard deviation	...	...	63.1	...	...	0.79	...	...	2.25	...	...	6.67
Standard deviation $\div\sqrt{8}$	...	...	63.1	...	...	0.79	...	...	2.25	...	...	6.67

It will he noticed that the kiln-dried seed gave on an average the larger yield. of corn and straw, but that the quality was almost always inferior. At first sight this might be supposed to be due to superior germinating power in the kiln-dried seed, but my farming friends tell me that the effect of this would be that the kiln-dried seed would produce the better quality barley. Dr Voelcker draws the conclusion: "In such seasons as 1899 and 1900 there is no particular advantage in kiln-drying before mowing." Our examination completely justifies this and adds "and the quality of the resulting barley is inferior though the yield may be greater."

In this case I propose to use the approximation given by the normal curve with standard deviation $s/\sqrt{n-3}$ and therefore use Sheppard's tables, looking up the difference divided by $S/\sqrt{8}$. The probability in the case of yield of corn per acre is given by looking up $33.7/22.3 = 1.51$ in Sheppard's tables. This gives $p = 0.934$, or the odds are about 14 to 1 that kiln-dried corn gives the higher yield.

Similarly $0.91/0.28 = 3.25$, corresponding to $p = 0.9994$,[2] so that the odds are very great that kiln-dried seed gives barley of a worse quality than seed which has not been kiln-dried.

Similarly, it is about 11 to 1 that kiln-dried seed gives more straw and about 2 to 1 that the total value of the crop is less with kiln-dried seed.

[2] As pointed out in Section V, the normal curve gives too large a value for p when the probability is large. I find the true value in this case to be $p = 0.9976$. It matters little, however, to a conclusion of this kind whether the odds in its favour are 1660 to 1 or merely 416 to 1.

Section X. Conclusions

1. A curve has been found representing the frequency distribution of standard deviations of samples drawn from a normal population.

2. A curve has been found representing the frequency distribution of the means of the such samples, when these values are measured from the mean of the population in terms of the standard deviation of the sample.

3. It has been shown that the curve represents the facts fairly well even when the distribution of the population is not strictly normal.

4. Tables are given by which it can be judged whether a series of experiments, however short, have given a result which conforms to any required standard of accuracy or whether it is necessary to continue the investigation.

Finally I should like to express my thanks to Prof. Karl Pearson, without whose constant advice and criticism this paper could not have been written.

[*Biometrika*, **6** (1908), pp. 1–25, reprinted on pp. 11–34 in *"Student's" Collected Papers*, Edited by E. S. Pearson and John Wishart with a Foreword by Launce McMullen, Cambridge University Press for the Biometrika Trustees, 1942.]