Onderstepoort Journal of Veterinary Science and Animal Industry, Volume 9, Number 1, July, 1937.

Printed in the Union of South Africa by the Government Printer, Pretoria.

Wool Studies.

II. The Frequency Distribution of Merino Wool Fibre Thickness Measurements.

By A. P. MALAN, Section of Statistics, Onderstepoort.

1. Introduction.

Practically the whole of the statistical theory and practice of modern agricultural and biological experimentation is based on what is known as the *normal theory*. This means that the usual tests of significance of statistical coefficients are based on the assumption that the coefficients are estimated from " random samples from a normal population ".

Although the existing tables for significance tests are mostly based on the assumption of a normal parent population, it is known that the applicability of these tests is not always confined to strictly normal populations. It was demonstrated with a practical example by Eden and Yates (1933) that the lack of normality did not violate the application of Fisher's z-tests in the analyses of variance. However, it is equally true that the normal theory may not be applied indiscriminately to all data. It is therefore necessary to study the nature of the observed distribution functions in different fields of experimental work.

In cases where the observed variate is definitely not normally distributed it is sometimes possible to substitute a function of the observed variate as the new variable quantity, which becomes normally distributed. So, for instance in cases where the observed standard deviations for different samples vary in proportion to the respective mean values, it is reasonable to use the logarithms of the observed values as the variate for the statistical analysis of the data.

2. Wool Fibre Thickness Measurements.

Fibre thickness is measured at the Onderstepoort Wool Laboratory by the micro-camera method. The sampling consists of zoning the original sample and from each of these a small portion is taken and combined into a single sample. This sample is cut along its whole length into a large number of small fragments which are then mixed in a beaker of ether. A portion of this mixture is taken at random, dried and mounted in Euparal on a slide from which the required number of thicknesses is determined by means of a Zeiss-Hegener micro-camera.

The above method allows for the variation in thickness between fibres and the variation in thickness along the length of the same fibre, but in an uncontrolled way, in the sense that some fibres may contribute more to the observed variation than others according to the respective numbers of fragments included in the observed values. Various other objections may be made against the above procedure, and the involved problem of wool sampling is receiving a thorough investigation. It is hoped to give a more detailed discussion of wool sampling technique in a future study.

Fibre thickness measurements are known to have a skew frequency distribution which is not normal. The actual nature of this distribution has, to the author's knowledge, never been discussed and it is the intention of this paper to apply the logarithmic transformation to observed thickness measurements. This transformation was suggested by the constancy of the coefficients of variability for the same fibre population.

In Study I, Malan, van Wyk and Botha (1935) considered amongst others, the variation in fibre thickness measurements over a period of three years. Consecutive measurements were made of shoulder samples from a marked area on the skin to ensure that they represented the same fibre population. A striking feature of the results was the constancy of the coefficients of variability obtained for different years, notwithstanding a considerable change in the mean values. In a paper on some characteristics which enter into the assessment of wool quality, and their estimation in the fleece, Wildman (1935) used the logarithms of fibre thickness " on account of the proportionate relation between the standard deviation and mean ". This transformation into logarithms would be justified if the logarithms become normally distributed.

3. The Logarithmic Distribution.

A variate x is said to be normally distributed when the frequency in an infinitesimal interval dx is proportional to df, where:

$$df = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2s^2}} dx \quad \ldots\ldots(1)$$

Hence if the variate (x) in (1) is considered as the natural logarithm of fibre thickness (t), $x = log_e (t)$, the distribution function of t is given by:

$$df = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} log_e^2 \left(\frac{t}{m}\right)} \frac{dt}{t} \quad \ldots\ldots(2)$$

In (2) the parameters are m and σ, where m is the geometrical mean of the fibre thickness measurements and σ is the root mean square deviation of $log_e t$ from the natural logarithm of the geometrical mean. Evidently, therefore, σ is a measure of mean squared deviation as a proportion of the mean and 100σ is a measure of compound percentage deviation from the mean. The coefficient, σ, will be referred to as the *coefficient of "relative" variability.*

The properties of the above function (2) have been considered by various authors but the required results for its application in this paper are again deduced and graphically illustrated.

This function, (2), has its maximum where:

$$t = m\ e^{-\sigma^2} \quad \text{......(3)}$$

and its points of inflexion at:

$$t = m\ e^{-\frac{3}{2}\sigma^2 \pm \sigma\sqrt{1+\frac{\sigma^2}{4}}}$$

The moments (M') about the origin, where t is *zero*, are given by:

$$M'_n = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty t^n\, e^{-\frac{1}{2\sigma^2}\ log_e^2\left(\frac{t}{m}\right)}\ \frac{dt}{t}$$, which by putting

$\frac{1}{\sigma\sqrt{2\pi}}\ log_e\left(\frac{t}{m}\right) = x$, is easily shown to give:

$$M'_n = m^n\, e^{\frac{n^2\sigma^2}{2}} \quad \text{......(4)}$$

From (4) the first two moments are obtained by putting $n = 1$ and 2 respectively:

$$M'_1 = m\, e^{\frac{\sigma^2}{2}} = a, \text{ the arithmetical mean} \quad \text{......(5)}$$

$$M'_2 = m^2\, e^{2\sigma^2} \quad \text{......(6)}$$

Hence the parameters m and σ, expressed in terms of the first two moments are:

$$m = \frac{M_1'^2}{\sqrt{M'_2}} \quad \text{......(7)}$$

$$e^{\sigma^2} = \frac{M_2'}{M_1'^2} \quad \text{or} \quad \sigma^2 = log_e\left(\frac{M_2'}{M_1'^2}\right) \quad \text{......(8)}$$

From the above moments about zero (M') given by (4), the moments about the arithmetical mean (M), are found to be:

$$M_n = m^n\, e^{\frac{1}{2}n\sigma^2} \sum_{i=1}^{n} \left\{ (-1)^i \binom{n}{i}\, e^{\frac{1}{2}(n-i)(n-i-1)\sigma^2} \right\} \quad \text{......(9)}$$

Likewise, having got the moments, the cumulants or semi-invariants may be obtained by means of the known relationships between them.

From (9) the first two moments about the arithmetical mean, $a = M'_1$ are:

$$M_1 = 0$$
$$M_2 = m^2 e^{\sigma^2} (e^{\sigma^2} - 1) = s^2 \quad \ldots\ldots (10)$$

where s^2 is the ordinary variance.

Substituting from (5), which gives the arithmetical mean, a, the above equation (10) becomes:—

$$s^2 = a^2 (e^{\sigma^2} - 1) \quad \ldots\ldots (11)$$

$$\frac{s^2}{a^2} = e^{\sigma^2} - 1$$

$$= \sigma^2 + \frac{\sigma^4}{2!} + \frac{\sigma^6}{3!} + - - - -$$

$\simeq \sigma^2$ to a first approximation, and $\simeq \sigma^2 (1 + \frac{\sigma^2}{4})^2$ to a second approximation.

From (11) the value of σ may be expressed in terms of s/a as follows:

$$\sigma^2 = \log_e \left(1 + \frac{s^2}{a^2}\right) \quad \ldots\ldots (12)$$

$$= \frac{s^2}{a^2} - \tfrac{1}{2}\frac{s^4}{a^4} + \tfrac{1}{3}\frac{s^6}{a^6} - - - -$$

$= \frac{s^2}{a^2}$ to a first approximation, and

$= \frac{s^2}{a^2}\left(1 - \frac{1}{4}\frac{s^2}{a^2}\right)^2$ to a second approximation.

The following table illustrates the accuracy of the second approximation for σ in terms of s/a where 100σ was called the *coefficient of relative variability* and $100\,s/a$ the ordinary *coefficient of variability*.

TABLE I.

s/a.	0·1	0·2	0·3	0·4	0·5
$\sigma = \sqrt{\log_e \left(1 + \frac{s^2}{a^2}\right)}$	0·099,751	0·198,040	0·293,560	0·385,253	0·472,380
$\frac{s}{a}\left(1 - \frac{1}{4}\frac{s^2}{a^2}\right)$	0·099,750	0·198,000	0·293,250	0·384,000	0·468,750
Difference..............	0·000,001	0·000,040	0·000,310	0·001,253	0·003,630

From the differences in the last row it is clear that for a coefficient of variability below 30 per cent. the second approximation for σ is sufficiently accurate for most purposes.

4. The Fitting of the Logarithmic Curve.

The statistical coefficients m and σ for any observed distribution may be estimated from the first two moments about zero by means of the relations (4) and (5) respectively. Once the estimates of m and σ are known the fitting of the theoretical curve (2) to the observed frequencies, is a matter of routine procedure. The expected frequency of values between *zero* and t is proportional to:

$$A_t = \frac{1}{\sigma\sqrt{2\pi}} \int_0^t e^{-\frac{1}{2\sigma^2} \log_e^2\left(\frac{t}{m}\right)} \frac{dt}{t} \quad \text{.........................(13)}$$

By putting $\frac{1}{\sigma} \log_e\left(\frac{t}{m}\right) = x$ the above integral takes the familiar form,

$$A_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}x^2}\, dx \quad \text{.........................(14)}$$

Therefore the area of the "tail" of the logarithmic curve from o to t is equal to the area of the "tail" of the normal error function from $-\infty$ to $m\,e^{\sigma x}$. Hence by putting $x = \frac{1}{\sigma} \log_e\left(\frac{t}{m}\right)$ the required values of x for the given function to enter, e.g. Table II of Pearsons' *Tables for Statisticians and Biometricians, Pt. I*, are obtained.

Thus, when the total observed frequency is equal to N, the "expected" frequency between t_1 and t_2 is given by:

$$A_{t_2} - A_{t_1} = \frac{N}{\sigma\sqrt{2\pi}} \left[\int_0^{t_2} e^{-\frac{1}{2\sigma^2}\log_e^2\left(\frac{t}{m}\right)} \frac{dt}{t} - \int_0^{t_1} e^{-\frac{1}{2\sigma^2}\log_e^2\left(\frac{t}{m}\right)} \frac{dt}{t} \right]$$

$$= \frac{N}{\sqrt{2\pi}} \left[\int_{-\infty}^{x_2} e^{-\frac{x^2}{2}}\, dx - \int_{-\infty}^{x_1} e^{-\frac{x^2}{2}}\, dx \right]$$

Where $x_1 = \frac{1}{\sigma} \log_e\left(\frac{t_1}{m}\right)$ and $x_2 = \frac{1}{\sigma} \log_e\left(\frac{t_2}{m}\right)$

By calculating the "expected" frequencies for each group interval, the agreement between "observed" and "expected" frequencies may be tested by means of the χ^2 test for "goodness of fit".

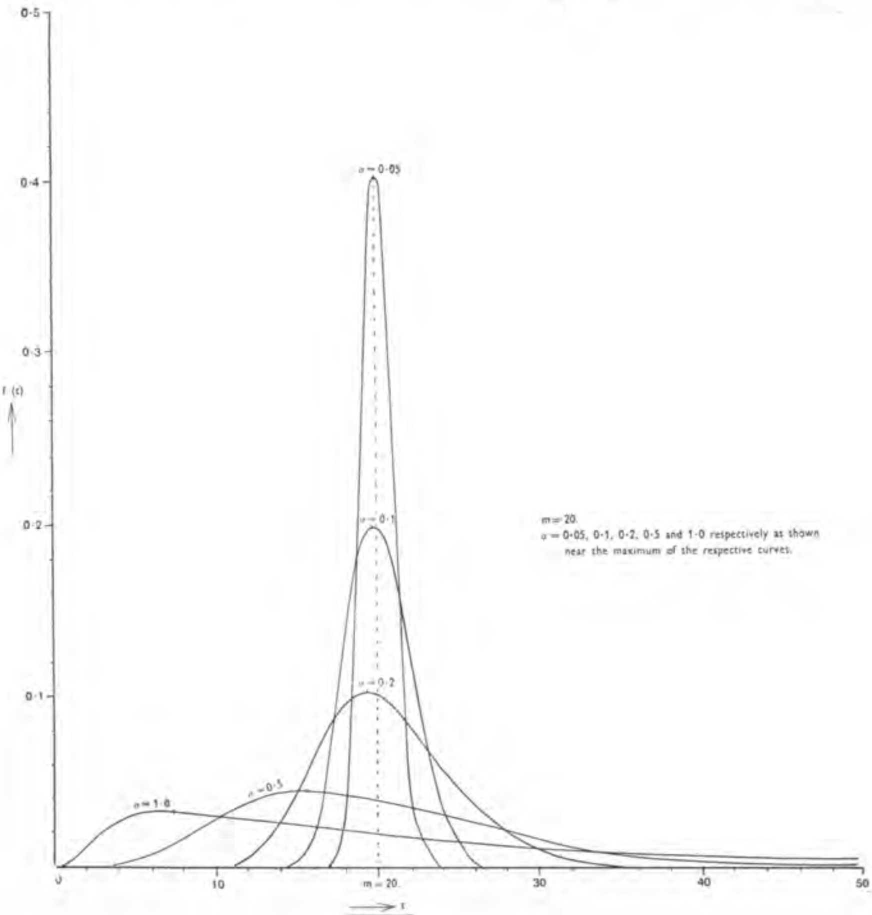
To draw the theoretical curve, estimated from the observed data, it is only necessary to note that the ordinate at any point t $(m\,e^{\sigma x})$ is obtained by multiplying the error function value $\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right)$ for $x \left(= \frac{1}{\sigma} \log_e\left(\frac{t}{m}\right)\right)$ by $\frac{N}{\sigma t}$.

5. The Shape of the Logarithmic Curve.

The normal function (1) has a familiar bell shaped form and is symmetrical about the mean (a), where the function has its maximum value. The variate x varies from $-\infty$ to $+\infty$ and the shape of the curve is flattened out with increasing values of the standard deviation s.

By the logarithmic transformation the function (2) becomes skew and the variate (t) remains positive between the limits $o \leq t \leq \infty$, with its maximum at the point $t = m\,e^{-\sigma^2}$ as given by (3). For the same value of the geometrical mean, the maximum point moves further away from the mean towards *zero* as σ, the *coefficient of relative deviation*, increases. The alteration in the shape of the logarithmic curve (2) for different values of σ is illustrated by *Chart A*, where the logarithmic curves are drawn with a constant geometrical mean $m = 20$ and *coefficients of relative deviation*, $\sigma = 0{\cdot}05$, $0{\cdot}1$, $0{\cdot}2$, $0{\cdot}5$ and $1{\cdot}0$ respectively. For these curves the maximum values are at $t = 19{\cdot}95$, $19{\cdot}80$, $19{\cdot}22$, $15{\cdot}58$ and $7{\cdot}36$ respectively. The lack of normality is clearly increased by increased values of σ,

Chart A.—The Logarithmic Function $f(t) = \dfrac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma^2}\log_e^2 \frac{t}{m}}\, \dfrac{1}{t}$



6. The Parameters of the Normal and Logarithmic Functions.

The normal function (1) is uniquely determined by the parameters a and s while the corresponding logarithmic function is determined by m and σ. The relationships between these parameters have been deduced and given in paragraph 3. It is of interest to consider the difference between these parameters when the two functions are applied to the same population. For this purpose populations have been selected with a constant arithmetical mean, (a), equal to 20 and standard deviations varying from 1 to 40, i.e. coefficients of variability from 5 per cent. to 200 per cent. These extreme values for the coefficients of variability are included because the shapes of the corresponding logarithmic curves depends only, according to (12), on the ratio of the standard deviation to the mean. The mean of 20 has been taken to represent more or less an average value for fibre thickness measurements. Obviously, when any other mean (a) is chosen with the same coefficient of variability the logarithmic deviation coefficients remain unaltered and specific values for the variates are obtained from the results given below by simply multiplying the given values by the ratio $^m/20$.

The position is clearly illustrated by Table I where the logarithmic values of the mean, maximum point and deviation coefficients are given for each coefficient of variability. This table also includes columns to show the differences between the means and between deviation coefficients. The increased skewness of the logarithmic curve for greater values of the deviation coefficients is further illustrated by the difference column between the geometrical mean (m) and the maximum point ($T\ max$).

Table II.

$a = 20$. s	s/a	m	T max	σ	$a-m$	$m -$ T max	100 x $(s/a-\sigma)$
1..................	0·05	19·975	19·925	0·04997	0·025	0·025	0·003
2..................	0·10	19·901	19·704	0·09975	0·099	0·197	0·025
3..................	0·15	19·779	19·343	0·14917	0·221	0·436	0·083
4..................	0·20	19·612	18·857	0·19804	0·388	0·755	0·196
5..................	0·25	19·403	18·262	0·24622	0·597	1·141	0·378
6..................	0·30	19·157	17·575	0·29354	0·843	1·582	0·646
8..................	0·40	18·570	16·008	0·38525	1·430	2·562	1·475
10..................	0·50	17·880	14·311	0·47234	2·111	3·578	2·766
12..................	0·60	17·150	12·610	0·55451	2·850	4·540	4·549
15..................	0·75	16·000	10·240	0·66805	4·000	5·760	8·195
20..................	1·0	14·144	7·072	0·83255	5·856	7·072	16·745
24..................	1·2	12·804	5·247	0·94446	7·196	7·557	25·554
30..................	1·5	11·094	3·414	1·0850	8·906	7·680	41·50
36..................	1·8	9·713	2·291	1·2019	10·287	7·422	59·81
40..................	2·0	8·944	1·789	1·2686	11·056	7·155	73·14

a = arithmetical mean,
s = normal standard deviation,
m = geometrical mean,
σ = logarithmic or ' relative ' deviation coefficient,
T max = point where the logarithmic curve has its maximum.

7. The Probability Integral when the Parameters of the Normal Population are used for a Logarithmic Distribution.

For the Normal Population given by (1), the area under the curve beyond a point $x=a+ns$ is given by $1 - P_{ns}$ where:

$$P_{ns} = \frac{1}{s\sqrt{2\pi}} \int_{-\infty}^{a+ns} e^{-\frac{(x-a)^2}{2s^2}} dx \quad \text{.......(15)}$$

The total area outside the limits $a \pm ns$ is given by $2\ (1 - P_{ns})$, which is equal to:—

$$\left\{ 1 - \frac{1}{s\sqrt{2\pi}} \int_{a-ns}^{a+ns} e^{-\frac{(x-a)^2}{2s^2}} dx \right\} \quad \text{.......(16)}$$

Using the limits $t=a \pm ns$ for the logarithmic population (2) the integral (15) becomes:

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a+ns} e^{-\frac{1}{2\sigma^2} \log_e^2 \left(\frac{t}{m}\right)} \frac{dt}{t} \quad \text{.......(17)}$$

where m and σ are given by (7) and (8), and the value of (16) is given by:

$$1 - \frac{1}{\sigma\sqrt{2\pi}} \int_{a-ns}^{a+ns} e^{-\frac{1}{2\sigma^2} \log_e^2 \left(\frac{t}{m}\right)} \frac{dt}{t} \quad \text{.......(18)}$$

Hence, to obtain the value of these integrals, it follows from paragraph (4) that the values of these limits by which Pearson's *Tables for Statisticians and Biometricians* is to be entered, are:

$$t = \frac{1}{\sigma} \log_e \left(\frac{a \pm ns}{m}\right)$$

These limits are affected by the arithmetical mean and the standard deviation, or to be more accurate, by the ordinary coefficient of variability. It has, therefore, been decided to consider all the values of s/a from 0·05 to 2. The probabilities of obtaining values of t below $(a - ns)$, beyond $(a+ns)$ and both below and beyond these two values of t, where $n=1$, 2 and 3 and the distribution of t are given by the logarithmic function (2). These probabilities for $n=1$, 2 and 3 are represented by charts B, C and D respectively. The probabilities of getting values of $t \leq a - ns$ and $t \geq a + ns$, written $P\{t \leq (a-ns)\} = P_1$ $P\{t \geq (a+ns)\} = P_2$, and also the values for $P_1 + P_2$, as s/a varies from 0 to 2, are shown by curves on these charts. Furthermore the total probabilities (P) of getting deviations from the mean greater than *one*, *two* or *three* times the standard deviation, σ, are shown by dotted lines on the respective charts. Hence the difference between the dotted line value and the value on the $P_1 + P_2$ line for a particular value of s/a show the difference between the assumed normal theory value and the actual value based on the logarithmic distribution.

CHART B.—Giving $P\{t \leq (a-s)\} = P_1$, $P\{t \geq (a+s)\} = P_2$ and $P_1 + P_2$ for $m = 20$ and $o \leq \frac{s}{a} \leq 2$.

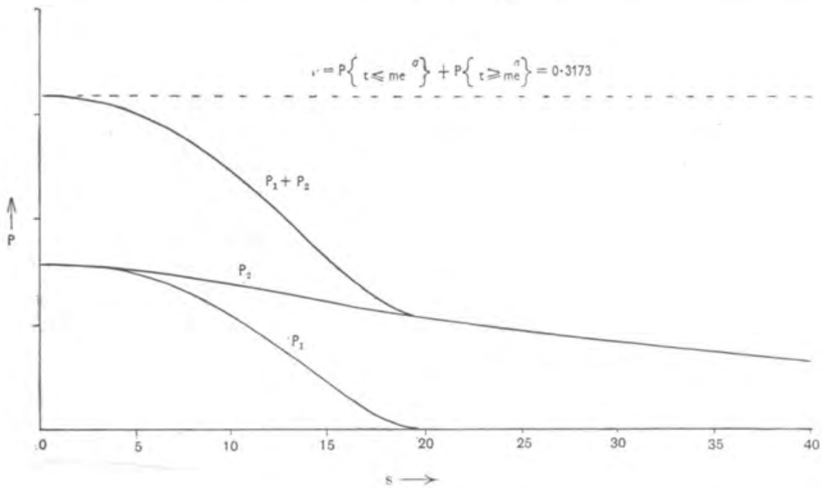


CHART C.—Giving $P\{t \leq (a-2s)\} = P_1$, $P\{t \geq (a+2s)\} = P_2$ and $P_1 + P_2$ for the values $m = 20$ and $o \leq \frac{s}{a} \leq 2$.

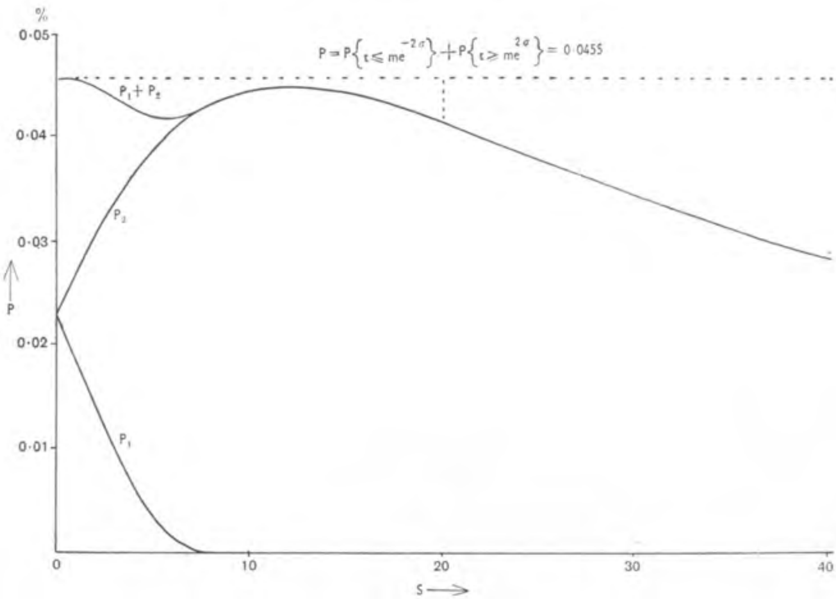
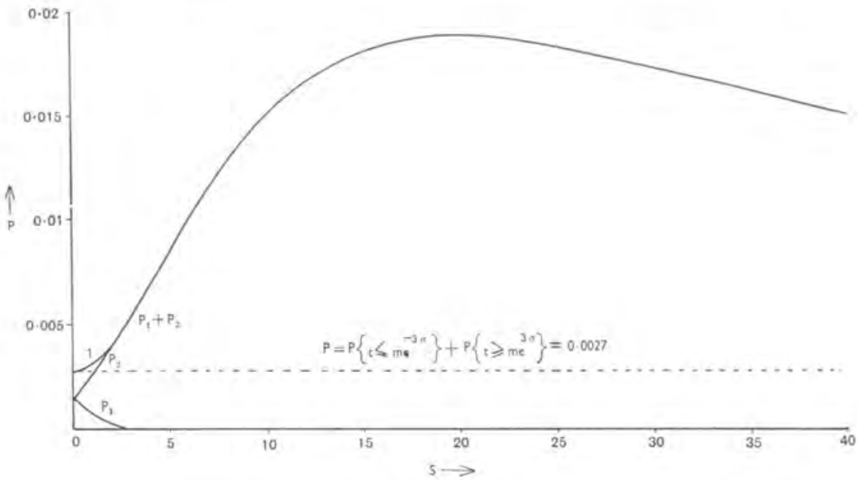


CHART D.—Giving $P\{t \leq (a-3s)\} = P_1$, $P\{t \geq (a+3s)\} = P_2$ and $P_1 + P_2$ for values of $m = 20$ and $o \leq \frac{s}{a} \leq 2$.



The two probability values become more equal as $s/_a$ tends to *zero*. In Chart B., i.e. where $n=1$, the normal theory value over-estimates the actual probability and this discrepancy increases with increasing values of $s/_a$. P_1 rapidly approaches zero and the two probabilities are only approximately equal in the neighbourhood of $s/_a = 0$, i.e. where the logarithmic distribution approaches the normal curve.

In Chart C, i.e. $n=2$, P_1 becomes zero very rapidly but P_1+P_2 remains approximately equal to P over a wide range of values of $s/_a$. The normal theory provides a good approximation for P, when the values of the coefficient of variability *($100 s/_a$)* are below *100 per cent.* This is very important from a practical point of view. The probability of getting deviations from the mean greater than twice the standard deviation is near the 5 per cent. value which forms a critical value in test criteria. Hence for $n=2$ no serious error will follow when the normal theory is applied to a logarithmic distribution to estimate P, provided $s/_a$ is below 1.

When $n=3$ it follows from Chart D that P_1 becomes *zero* when $s/_a$ is still extremely small whereas P_2 increases very rapidly with increasing values of $s/_a$ and at reasonably small values of $s/_a$, P_2 becomes many times greater than the normal theory value $P = 0{\cdot}0027$. The discrepancy between the normal theory value P and the probability for the logarithmic distribution is such that for reasonable values of $s/_a$, $P = 0{\cdot}0027$ considerably underestimates the actual probability P_1+P_2, as calculated from the logarithmic curve.

The skewness of the logarithmic curve is demonstrated by the P_1 and P_2 curves on the above charts. P_1 rapidly tends to zero as n is increased and becomes zero when ns is equal to or greater than a.

8. Material.

The data in the present study, which intends the application of the logarithmic function (2) to observed frequency distributions in fibre thickness measurements, were obtained by two different sampling methods. Tht first group of samples, Group A, was taken by the method of practice at the Onderstepoort Wool Laboratory. This method is described in a previous paragraph. The two samples of Group B were obtained by mounting a small sample of stretched fibres on a slide for measurement.

The samples of Group A figure in an independent investigation but were placed at the author's disposal for the purpose of this study. There were altogether six samples from each of a fine, medium and a strong wool. These wool classes were selected from a large number of fleeces and ultimately sampled.

In the second group of samples every fibre was measured once only, in order to eliminate variation "within" fibres. The thicknesses in this group were also measured on an anti-logarithmic scale of which the scale divisions are proportional to numbers with equally spaced natural logarithms (Chart E). Thus when the histogram which represents observed frequencies is drawn on a logarithmic base the group intervals are equal and the logarithmic function (2) becomes normal. An image of the magnified scale is given below. (Chart E.)

9. Presentation of Data.

Group A.

This group of observations was taken from three different thickness classes and is presented accordingly in three separate tables. Together with each observed sample are given in adjacent columns the expected frequencies from both a normal and a logarithmic population, with the estimated mean and deviation coefficients below the respective columns. For the normal theory the standard deviations and coefficients of variability are both given but for the logarithmic distributions only the measure of percentage deviation is shown. The values of the ordinary coefficients of variability are about 20 per cent. and therefore approximately 0·2 per cent. greater than the coefficients of relative variability as shown by the tabulated values for $s/_a$ and σ in paragraph 3, Table I.

The agreement between "observed" and "expected" frequencies is measured by χ^2. The relative merits of the two distributions, normal and logarithmic, as the parent population, may be judged from the respective probabilities. These probabilities, given in the last row of the tables, refer to the chances of obtaining the samples from the respective theoretical distributions.

The logarithmic nature of the frequency distributions of fibre thickness measurements is further illustrated by Figures 1, 2 and 3, where the frequency histograms of the data in Tables III, IV and V respectively and the best fitting normal and logarithmic curves are produced. The observed distributions in these graphs are typical of observed distributions of fibre thickness measurements.

CHART E.—The anti-logarithmic Scale.

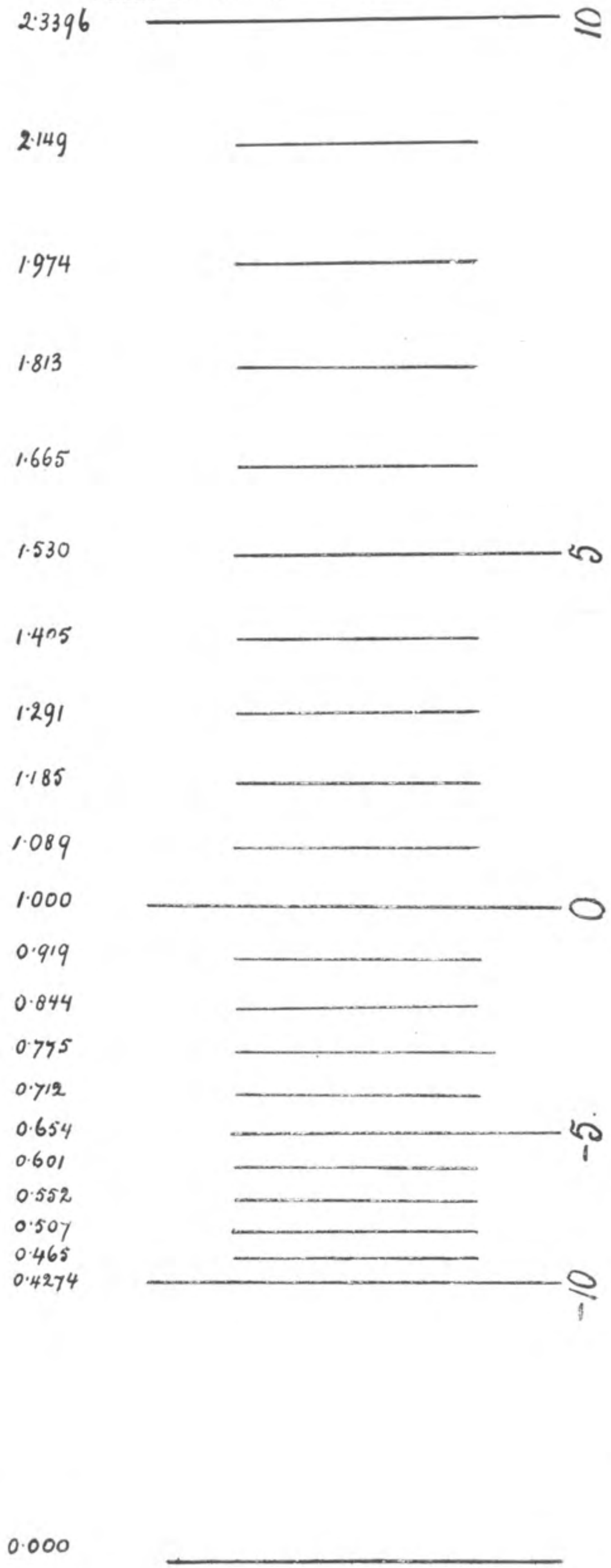


Table III.

Frequency Distribution in Fibre Thickness of a Fine Wool.

SAMPLES.

Group Interval. (μ).	17			18			19			30			31			32		
	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.
7·5—10..........	2	5·7	1·9	5	6·4	—	1	—	—	3	6·4	—	4	7·2	—	1	—	—
10 —12·5........	13	19·3	12·6	14	22·2	18·2	15	27·4	17·3	19	22·4	17·3	18	23·5	20·0	13	24·3	14·5
12·5—15..........	56	53·1	62·1	63	60·5	72·0	69	60·1	71·4	63	61·2	69·0	66	61·4	73·8	68	57·8	67·5
15 —17·5........	106	98·8	117·3	119	108·5	126·2	136	109·3	126·9	130	109·5	124·6	126	107·4	125·1	117	109·9	128·0
17·5—20..........	151	124·2	125·4	150	128·2	125·1	126	129·5	125·9	125	129·3	126·1	127	125·7	122·4	145	133·8	131·0
20 —22·5........	91	105·9	91·1	84	100·7	84·8	80	100·6	84·9	92	100·2	87·0	89	98·8	83·0	95	104·4	88·5
22·5—25..........	49	61·3	50·9	39	52·1	44·3	44	50·9	43·6	42	46·3	46·2	36	51·8	43·7	35	52·2	44·9
25 —27·5........	21	24·0	23·6	15	17·8	19·2	17	16·9	18·6	19	22·2	20·2	23	18·2	19·3	17	16·7	18·7
27·5—30..........	4	7·7	15·0	7	4·7	7·3	6	4·2	7·0	7	4·4	11·8	7	5·0	7·6	10	3·9	10
30 —32·5........	2	—	—	3	—	3·3	—	—	3·5	1	—	—	2	—	4·1	1	—	—
32·5—35..........	2	—	—	2	—	—	4	—	—	1	—	—	—	—	—	1	—	—
35 —37·5........	3	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—
37·5—40..........	—	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—
Total..............	500	500·0	499·9	501	501·1	500·4	499	498·9	499·1	502	501·9	502·2	499	499·0	499·0	503	503·0	503·1
Mean..............	—	19·0	18·6	—	18·5	18·1	—	18·5	18·1	—	18·5	18·3	—	18·5	18·1	—	18·6	18·2
S.D................	—	3·940	—	—	3·813	—	—	3·762	—	—	3·796	—	—	3·880	—	—	3·676	—
Percentage Deviation	—	20·8	20·5	—	20·6	20·4	—	20·3	20·1	—	20·5	20·3	—	21·0	20·8	—	19·8	19·6
χ^2.................	—	17·3	8·6	—	30·0	9·0	—	22·8	2·1	—	13·3	3·4	—	15·3	4·06	—	30·9	5·7
Degrees of freedom....	—	6	6	—	6	6	—	5	6	—	6	5	—	6	6	—	5	5
$P(\chi^2)$..............	—	0·008	0·20	—	—	0·17	—	—	0·91	—	0·039	0·64	—	0·018	0·67	—	—	0·34

Table IV.

Frequency Distributions in Fibre Thickness of a Medium Wool.

SAMPLES.

Group Interval (μ).	4			6			7			8			9			10		
	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.
10 —12·5.........	4	13·7	5·2	4	17·0	6·8	4	12·1	4·1	6	15·0	5·3	—	—	—	6	16·4	6·4
12·5—15..........	27	33·9	33·6	33	35·5	37·2	33	23·0	29·4	26	30·7	30·6	25	46·3	37·2	32	34·1	35·4
15 —17·5.........	94	73·5	90·4	82	71·4	88·2	69	67·4	81·9	74	62·6	77·4	88	66·1	81·3	80	68·9	85·2
17·5—20..........	129	111·1	144·2	136	104·8	117·5	127	105·6	119·6	119	94·4	109·9	137	100·5	114·3	130	102·0	115·6
20 —22·5.........	121	117·1	88·4	115	111·9	106·3	116	119·4	112·3	106	108·4	106·2	99	110·7	106·9	106	110·9	129·0
22·5—25..........	64	86·1	71·9	66	87·4	72·8	78	91·8	77·7	85	91·2	77·8	68	89·5	75·1	72	87·9	50·5
25 —27·5.........	33	44·1	38·0	42	49·6	41·4	46	50·8	42·9	40	57·4	46·9	43	52·6	44·2	40	51·2	42·4
27·5—30..........	14	15·8	17·2	10	20·6	20·4	17	24·9	20·3	22	27·0	24·7	20	22·5	21·6	19	21·8	21·3
30 —32·5.........	9	4·7	7·0	10	7·8	15·4	5	7·2	8·6	10	9·4	11·7	11	8·9	9·8	9	8·7	9·6
32·5—35..........	2	—	4·0	4	—	—	5	—	3·3	5	3·1	5·1	1	—	6·6	5	—	4·0
35 —37·5.........	2	—	—	1	—	—	1	—	1·8	5	—	3·5	3	—	—	2	—	2·6
37·5—40..........	1	—	—	2	—	—	1	—	—	1	—	—	—	—	—	—	—	—
40 —42·5.........	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	—	—
42·5—45..........	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—	—	—	—
45 —47·5.........	—	—	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—
Total..............	500	500·0	499·9	506	506·0	506·0	502	502·2	501·9	499	499·2	499·1	497	497·1	497·0	502	501·9	502·0
Mean...............	—	20·4	20·0	—	20·5	20·1	—	20·7	20·3	—	21·6	20·6	—	20·8	20·3	—	20·7	20·2
S.D................	—	4·100	—	—	4·389	—	—	4·178	—	—	4·557	—	—	4·387	—	—	4·424	—
Percentage Deviation	—	20·1	19·9	—	21·4	21·1	—	20·1	20·0	—	21·6	21·4	—	21·1	20·9	—	21·4	21·2
$\varkappa^2$................	—	44·1	18·3	—	42·8	11·8	—	19·3	6·1	—	41·5	5·7	—	39·1	10·7	—	34·0	16·5
Degrees of freedom..	—	6	7	—	6	6	—	6	8	—	7	8	—	5	6	—	6	8
P($\varkappa^2$)..............	—	—	0·011	—	—	0·068	—	0·004	0·64	—	—	0·67	—	—	0·10	—	—	0·036

TABLE V.

Frequency Distribution in Fibre Thickness of a Strong Wool.

SAMPLES.

Group Interval (μ).	20			22			33			34			36			37		
	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.	Observed.	Normal.	Logarithm.
7·5—10	1	—	—	—	—	—	1	—	—	1	—	—	4	2·9	—	—	—	—
10—12·5	4	17·9	6·2	2	—	3·4	2	9·2	2·3	6	21·9	7·1	6	9·6	—	3	11·1	2·6
12·5—15	25	28·9	29·1	17	32·0	17·2	14	21·3	18·4	42	42·1	40·6	15	28·6	28·6	16	26·8	22·3
15—17·5	71	54·0	67·1	61	49·3	58·6	57	48·4	59·2	87	84·5	104·6	69	64·7	77·4	71	61·6	73·1
17·5—20	107	80·4	97·9	96	82·4	99·3	102	82·2	100·9	173	130·3	154·7	138	106·1	126·0	135	104·7	125·0
20—22·5	87	94·6	98·5	110	104·0	109·9	109	104·9	110·9	168	153·9	156·7	143	130·3	133·6	143	131·8	136·6
22·5—25	84	89·4	79·5	92	99·7	89·8	96	100·7	89·6	110	139·5	121·8	94	118·2	104·4	99	122·7	109·5
25—27·5	58	66·9	54·2	66	71·9	56·8	53	72·3	58·5	69	96·4	78·5	66	79·9	65·6	77	84·5	69·7
27·5—30	27	40·0	32·6	30	39·0	35·4	38	39·1	32·3	54	51·2	44·8	33	39·9	35·2	34	43·1	37·7
30—32·5	20	18·9	18·1	17	16·0	16·9	16	15·8	15·9	23	20·8	21·6	22	14·9	16·9	20	16·2	23·1
32·5—35	8	9·9	9·3	3	6·3	14·7	8	6·1	7·1	9	8·4	10·5	6	5·0	7·4	6	5·6	8·2
35—37·5	5	—	4·6	3	—	—	2	—	5·0	3	—	8·1	2	—	5·0	1	—	—
37·5—40	1	—	4·0	3	—	—	1	—	—	2	—	—	1	—	—	1	—	—
40—42·5	2	—	—	—	—	—	1	—	—	2	—	—	1	—	—	—	—	—
42·5—45	1	—	—	1	—	—	—	—	—	—	—	—	—	—	—	1	—	—
45—47·5	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	—	—
Total	501	500·9	501·1	501	501·0	502·0	500	500·0	500·0	749	749·0	749·0	600	600·1	600·1	608	608·1	607·8
Mean	—	21·8	21·2	—	22·1	21·6	—	22·1	21·7	—	21·6	21·0	—	21·7	21·2	—	21·9	21·4
S.D.	—	5·182	—	—	4·654	—	—	4·610	—	—	4·786	—	—	4·516	—	—	4·496	—
Percentage Deviation	—	23·7	23·4	—	21·1	20·8	—	20·8	20·6	—	22·2	21·9	—	20·8	20·6	—	20·5	20·3
$\varkappa^2$	—	35·5	5·2	—	16·0	4·75	—	26·2	3·7	—	47·2	11·0	—	36·3	6·3	—	32·9	5·77
Degree of freedom	—	7	9	—	6	7	—	7	8	—	7	8	—	8	7	—	7	7
P($\varkappa^2$)	—	—	0·82	—	0·014	0·69	—	—	0·88	—	—	0·20	—	—	0·51	—	—	0·59

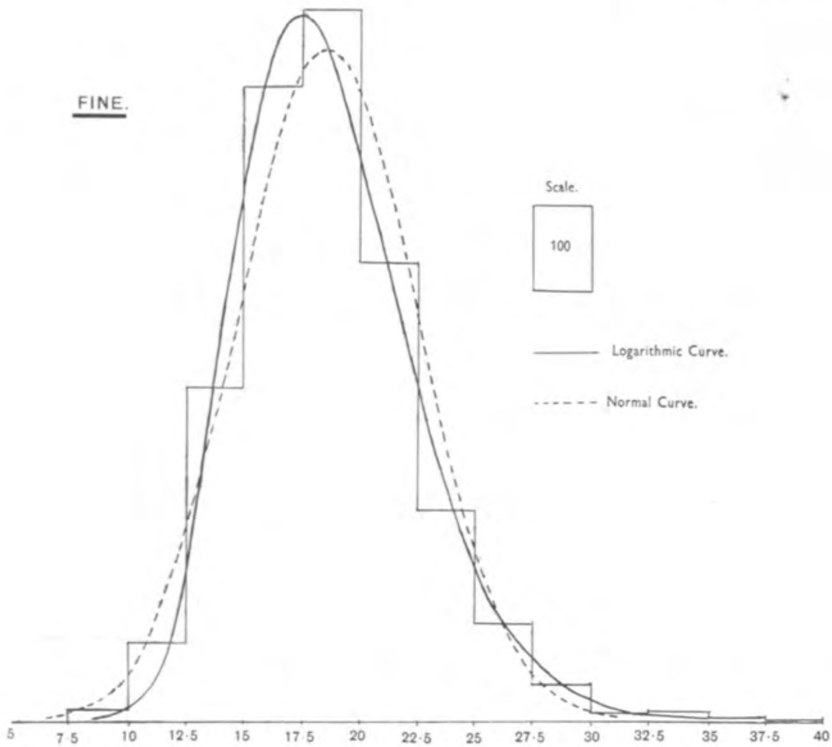


FIG. 1.—The Frequency Distribution of 3005 Fibre Thickness Measurements of a Fine Wool (μ).

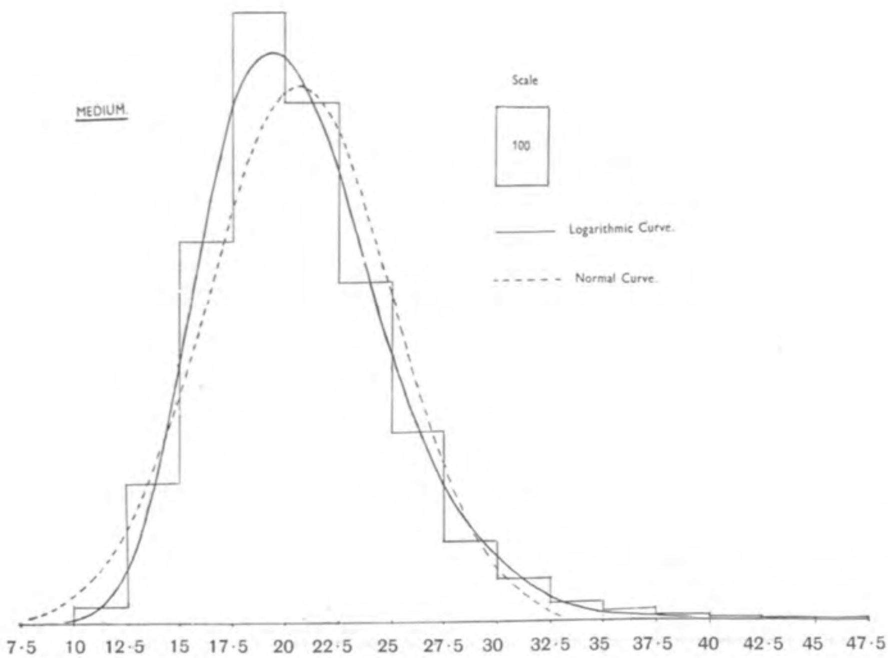


FIG. 2.—The Frequency Distribution of 3006 Fibre Thickness Measurements of a Medium Wool (μ).

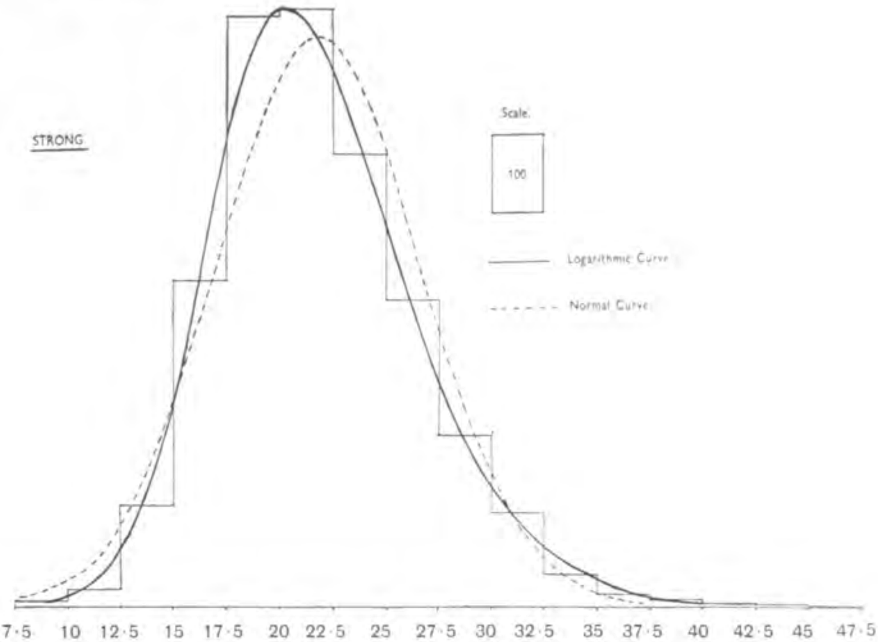


FIG. 3.—The Frequency Distribution of 3459 Fibre Thickness Measurements of a Strong Wool(*a*).

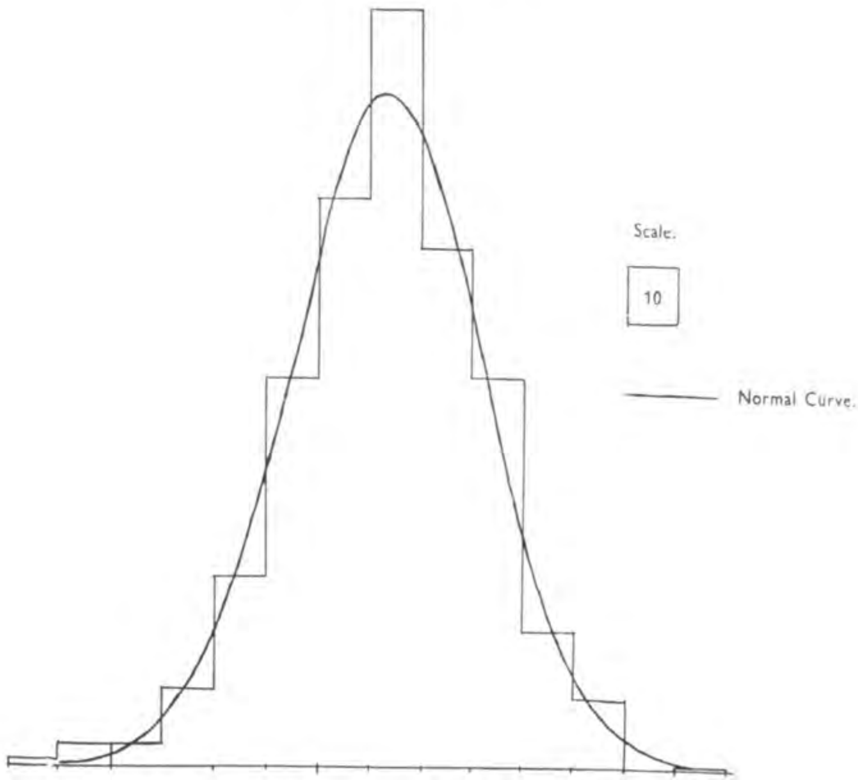


FIG. 4.—The Distribution of the Natural Logarithms of 557 Fibre Thickness Measurements.

Group B.

The two samples in this group were obtained from slides of stretched fibres, representing an ordinary shoulder sample and a sample of tops respectively. These measurements were taken by two measuring scales, the ordinary scale with a group interval of $2 \cdot 5\ \mu$ [Table VI (*a*)], and an anti-logarithmic scale (Chart E) with a difference of $0 \cdot 085$ between the natural logarithms of consecutive divisions [Table VI (*b*)]. Unit distance on this latter scale was adjusted at approximately $20 \cdot 5\ \mu$ for the shoulder sample and $21\ \mu$ for the tops. This adjustment of the magnification to obtain the precise value in μ which corresponds with unity on the anti-logarithmic scale, needs great care. Measurements by the ordinary scale in the eyepiece are not affected by alterations in magnification but for a loose scale the degree of magnification is essential for the ultimate conversion of coefficients into units of μ.

The logarithmic distribution of the ordinary measurements and the normality of the logarithmic measurements are clearly shown by Table VI. The normality is further illustrated in the cases of the shoulder sample, by Fig. 4, where the observed frequencies and best fitting normal curve for the logarithmic measurements for the shoulder sample are shown.

10. Discussion.

When the normal theory is considered the apparent deficient number of relatively thin fibres and excessive number of thick fibres are obvious. This is shown by all samples and illustrated by the graphs in Figures 1, 2 and 3. The observed distributions are decidedly skew and on the whole by no means normal. The probabilities for χ^2 in the case of the normal curve are almost throughout highly significant and in the majority of cases so small, less than $0 \cdot 001$, that the values are not given in the tables. When the limit for significance is taken at the one per cent. probability level it is seen that there are only two samples, (3) and (31) in Table III, none in Table IV, one (22) in Table V, and one (the shoulder sample) in Table VI (*a*), which can reasonably be assumed as random samples from a normal population. When the 5 per cent. probability level is taken not a single sample can be assumed to come from a normal population.

Considering, however, the assumption that these samples are taken from logarithmic populations the position is reversed and the agreements are on the whole within the ranges of reasonable expectation. The general trend of the logarithmic curve closely follows that of the observed histogram as shown by Figures 1, 2 and 3. The probabilities for χ^2 vary between $0 \cdot 91$ (Table III, sample 19) and $0 \cdot 011$ (Table IV,, sample 4). None of the samples disagree significantly with the logarithmic theory when the one per cent. probability level is assumed. For a 5 per cent. probability level there are only three samples (4, 6 and 10) in Table IV, and none in the other tables for which the χ^2 values are significant. In this connection it is interesting to note that when modal frequency is grouped with the

TABLE VI.

Frequency Distribution in Fibre Thickness.

Group Interval (μ).	(a) Ordinary Scale. (1) Shoulder Sample. Observed.	Normal.	Logarithm.	(2) Tops. Observed.	Normal.	Logarithm.	Group Interval. (Mid. Pts.)	(b) Logarithmic Scale. (1) Shoulder Sample. Observed.	Normal.	(2) Tops. Observed.	Normal.
10 —12·5	—	—	—	1	—	—	−6·5	—	—	2	—
12·5—15	1	—	—	10	16·3	9·4	−5·5	—	—	—	—
15 —17·5	4	9·4	4·5	27	33·0	37·0	−4·5	1	—	9	—
17·5—20	29	29·5	30·7	86	62·3	75·0	−3·5	4	—	8	20·2
20 —22·5	91	74·1	89·6	93	85·1	89·0	−2·5	4	5·7	27	25·4
22·5—25	125	124·3	138·1	75	84·1	78·9	−1·5	14	15·1	35	42·1
25 —27·5	151	139·6	132·1	40	60·5	50·3	−0·5	34	38·6	64	57·6
27·5—30	85	105·0	88·7	36	31·0	27·0	+0·5	69·5	74·1	70	70·2
30 —32·5	47	52·6	45·3	14	11·6	12·6	1·5	101·5	107·9	55	61·0
32·5—35	16	17·7	18·7	3	3·8	8·5	2·5	135·5	119·0	49	47·0
35 —37·5	6	4·6	9·4	2	—	—	3·5	92·5	99·3	26	29·8
37·5—40	1	—	—	1	—	—	4·5	70	62·8	20	15·7
40 —42·5	—	—	—	—	—	—	5·5	25	29·9	7	10·2
42·5—45	1	—	—	—	—	—	6·5	13	14·5	1	—
	—	—	—	—	—	—	7·5	—	—	1	—
	—	—	—	—	—	—	8·5	1	—	—	—
Total	557	556·8	557·1	388	388·3	387·7	—	567	566·9	374	373·9
Mean	—	25·57	25·27	—	22·4	22·0	—	—	2·35	—	0·642
S.D.	—	3·87	—	—	4·33	—	—	—	3·504	—	5·145
Percentage deviation	—	15·2	15·0	—	19·3	19·1	—	—	—	—	—
χ^2	—	14·9	5·3	—	23·4	10·8	—	—	7·6	—	5·8
Degree of freedom	—	6	6	—	6	6	—	—	7	—	7
$P(\chi^2)$	—	0·021	0·51	—	—	0·10	—	—	0·37	—	0·56

one following it, the significance is removed in all three cases. So for instance the probabilities for samples 4 and 10 (Table IV) become 0·42 and 0·85 respectively.

The distributions of Table VI (a) are in close agreement with those in the first three tables in that the normal theory does not seem to account for the observed distributions while the logarithmic theory supplies a good " fit " for both samples. Both methods of sampling, therefore, gave distributions which conform with the logarithmic function. This fact is further illustrated by the distributions in Table VI (*b*). The latter observations were obtained by using the anti-logarithmic scale as a measuring rule. The χ^2 values for these samples are in agreement with the assumption that they are taken from a population which becomes normally distributed when the logarithms of the values are taken.

In view of all these samples it appears reasonable to take the logarithms of wool fibre diameter measurements for purposes of statistical analysis. When this is done the arithmetical mean is replaced by the geometrical mean and since the latter is for fibre thickness measurements always slightly less than the former, it is advisable to give the quality numbers and class standards in terms of both means. The geometrical mean has various advantages for this purpose, being nearer the modal value and equal to the median. When the arithmetical mean is used to classify wool it is possible that the greater majority of fibres may actually fall in a lower class, which is obviated by the use of the geometrical mean.

11. Summary.

The logarithmic nature of distributions in wool fibre thickness measurements has been suggested by the constancy of the coefficient of variability in previous work.

The distribution of a variable, the logarithm of which is normally distributed, is discussed.

The application of what is in the text called the logarithmic function to 18 different samples is given and the " fit " compared with that of the normal distribution.

Two further samples, which were also measured by an anti-logarithmic scale, are included and show that the logarithms of fibre thicknesses are normally distributed.

The logarithmic nature of the distributions of fibre thickness measurements and the normality of the logarithms of such measurements are illustrated by Figures 1, 2, 3 and 4 respectively.

It is suggested that the logarithms of fibre thickness measurements be used for statistical analysis. This would mean that the arithmetical mean is to be replaced by the geometrical mean to represent average fibre thickness.

12. Acknowledgments.

The Author wishes to express his indebtedness to Dr. V. Bosman of the Wool Section for permission to use the data given in Tables III, IV and V and to Mr. C. M. van Wyk for kindly supplying the said data, which were taken in the Wool Section under his supervision, and to Mr. D. F. du Toit who assisted in the computations.

References.

EDEN, T., and YATES, F. (1933). On the validity of Fisher's Z-test when applied to an actual example of non-normal data. *Jnl. Agric. Sci.*, Vol. 23, pp. 6-16.

MALAN, A. P., VAN WYK, C. M., and BOTHA, M. L. (1935). The variation and interdependence of merino fleece and fibre characteristics. *Onderstepoort Jnl. Vet. Sc. and An. Ind.*, Vol. 5, No. 2, pp. 519-530.

WILDMAN, A. B. (1936). Some characteristics which enter into the assessment of wool quality and their estimation in the fleece. *Jnl. Text. Inst.*, Vol. 27, No. 7, pp. 181-192.