THE THEORY AND APPLICATION OF TRANSFORMATION IN STATISTICS

by

Anis Ismail Kanjo

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

December 1962

Blacksburg, Virginia

TABLE OF CONTENTS

Table of Contents (cont'd.)

Table of Contents (cont'd.)

Table of Contents (cont'd.)

Table of Contents (cont'd.)

Table of Contents (cont'd.)

INTRODUCTION

In order that the analysis of variance technique be adequately applicable, we have to consider certain assumptions such as randomness, additivity, homogeneity, normality, and uncorrelated errors. The reliability of the analysis depends upon the degree of fulfillment of these assumptions.

But in practice we frequently face the situation where the homogeneity requirement is violated by the nature of the data. For instance, we cannot claim homogeneity in data governed by a distribution whose variance is a function of the mean. In fact this is, generally, the case for any distribution depending on a single parameter. Two cases of interest are the Poisson distribution ($f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$) in which $E(x) = Var(x) = \lambda$ and the Binomial distribution ($f(x) = \binom{n}{x} p^x q^{n-x}$) in which $E(x) = np$ and $Var(x) = npq$.

It seems natural in such cases to seek transformations which make the variance independent of the mean, i.e., these transformations stabilize the variance. Examples of variance stabilization appear in the sequel in Part I.

Variate transformations are also useful in the following problems:

1. The so-called tail problem which relates to the calculation of the probability that a variate exceeds or is less than a given constant value.

2. The confidence problem which relates to the numerical determination of confidence limits for an unknown parameter.

In some types of distributions, particularly those which involve a discrete variate, these problems may involve lengthy computations. The situation may be alleviated by resorting to a well-chosen transformation which more or less produces a normally distributed variate. In this way percentage points and confidence intervals may be set up (approximately) by using the normal probability integral. Cases of this sort are considered in some detail in Part II.

In Part III we discuss two special types of transformations, namely the probability integral transformation (sometimes described as probits) and the logarithmic transformation (sometimes described as logits).

PART I

This is divided into two sections, A and B. In Section A we consider underlying assumptions and practical aspects of transformations in the analysis of variance. In Section B, theoretical considerations relating to transformations are discussed.

1. Underlying assumptions and practical aspects of transformations in the analysis of variance

1.1 Assumptions: One of the purposes of the analysis of variance is testing hypotheses which are frequently linear in form. Another purpose is the estimation of components of variance. In both these cases there are certain underlying assumptions which have to be made. To fix ideas we shall consider these assumptions in relation to a simple analysis of variance model in which the parameters of interest are the population means (model I).

Assumption 1: (Randomness) The observations consist of random variables that are distributed about fixed true mean values.

Consider for example a rectangular array of observations $[x_{ij}]$ where $i = 1, 2 \ldots r$, $j = 1, 2 \ldots c$. From the assumption we have $E(x_{ij}) = m_{ij}$ and thus the expectation of any linear function of the x_{ij}'s is equal to the same linear function of the m_{ij}'s.

Assumption 2: (Additivity) The mean value of any observation is a linear function of a certain set of parameters, i.e., we may express x_{ij} in the form:

$$x_{ij} = \mu + a_i + b_j + \varepsilon_{ij}$$

where μ, a_i and b_j are fixed and ε_{ij} is a random variate.

Assumption 3: (Equal variances and zero correlations) The random component of any observation (ε_{ij}) has the same constant variance, and the covariance between any two observations is zero.

Assumption 4: (Normality) The observations are jointly distributed in a multivariate normal form. This assumption in conjunction with that of zero covariances implies the mutual independence of the observations. Using these assumptions we can compute the variance of any linear function of the observations easily and carry out an exact test of significance.

Summarizing, we say that the necessary and sufficient conditions for the strict validity of the analysis of variance procedure for solving problems of model I with respect to data arranged in a rectangular array, are that:

$$x_{ij} = \mu + a_i + b_j + \varepsilon_{ij} \quad i=1,2...r, \; j=1,2...c \quad .$$

where the a_i, b_j, and μ are constants and the ε_{ij} are normally and independently distributed as $N(0, \sigma^2)$.

1.2 Randomized block design with additivity:

1.2.1 Model and randomization procedure:
We shall discuss an example of a randomized block design from the finite model point of view: that is, we shall assume that our observations are taken from a finite population. We shall denote the yield with treatment k (k = 1, 2 ... t) on plot j (j = 1, 2 ... t) of block i (i = 1,2 ...) by y_{ijk}, and we then have the identity:

$$y_{ijk}=y_{...}+(y_{i..}-y_{...})+(y_{ijk}-y_{ij.})+(y_{ij.}-y_{i..}) \quad . \tag{1}$$

Assuming that we have additive treatment effects we may write:

$$y_{ijk} - y_{ij.} = t_k \tag{2}$$

for all i and j. The identity (1) now becomes:

$$y_{ijk} = \mu + b_i + t_k + e_{ij} \tag{3}$$

where $\mu = y_{..}$, $b_i = y_{i..} - y_{...}$, $e_{ij} = y_{ij.} - y_{i..}$. In fact we observe the yield of treatment k on a randomly chosen plot in the block. Denoting the observed yield of treatment k in block i by y_{ik}, we may write:

$$y_{ik} = \mu + b_i + t_k + \sum_j \delta_{ij}^k e_{ij} \tag{4}$$

where

$\delta^k_{ij} = 1$ if treatment k occurs on plot j in the ith block.

$\delta^k_{ij} = 0$ otherwise.

The random error attached to any observed yield is the whole expression $\sum_j \delta^k_{ij} e_{ij}$ in which any particular e_{ij} is a fixed variable which we do not know. Moreover, δ^k_{ij} is a random variable whose distribution is determined by the randomization procedure used in obtaining the particular experimental plan. The properties of δ^k_{ij} are as follows:

1. Prob. $(\delta^k_{ij} = 1) = \frac{1}{t}$, for any i, j, k .
2. $\delta^k_{ij'} = 0$, given that $\delta^k_{ij} = 1$, for all $j' \neq j$.
3. $\delta^{k'}_{ij} = 0$, given that $\delta^k_{ij} = 1$, for all $k' \neq k$.
4. δ^k_{ij} and $\delta^{k'}_{i'j'}$ are independent if $i' \neq i$ for any j, j', k, k' .
5. Prob.$(\delta^{k'}_{ij'} = 1 / \delta^k_{ij} = 1) = \frac{1}{t-1}$ for $j' \neq j$, $k' \neq k$ $(t \geq 2)$.

Now from (4):

$$T_k = \sum_{i=1}^{r} y_{ik} = r\mu + rt_k + \sum_{ij}\sum \delta^k_{ij} e_{ij} \qquad (5)$$

and moreover $\sum_j e_{ij} = 0$ and $E(\delta^k_{ij}) = \frac{1}{t}$.

Hence:

$$E(T_k) = r\mu + rt_k + \frac{1}{t} \sum_{ij}\sum e_{ij} = r\mu + rt_k \quad . \qquad (6)$$

Now:

$$\text{Var}(T_k) = E(\sum_{ij}\sum \delta^k_{ij} e_{ij})^2$$

and:

$$(\sum_{ij}\sum \delta^k_{ij} e_{ij})^2 = \sum_{ij}\sum (\delta^k_{ij})^2 e^2_{ij} + \sum_i \sum_{j'\neq j} \delta^k_{ij} \delta^k_{ij'} e_{ij} e_{ij'}$$

$$+ \sum_{i'\neq i} \sum_j \sum_{j'} \delta^k_{ij} \delta^k_{i'j'} e_{ij} e_{i'j'} .$$

But:

$$E(\delta^k_{ij})^2 = \frac{1}{t} .$$

$$E(\delta^k_{ij} \delta^k_{ij'}) = 0 , \quad j \neq j' .$$

$$E(\delta^k_{ij} \delta^k_{i'j'}) = \frac{1}{t^2} , \quad i' \neq i \text{ for any } j, j' .$$

Hence:

$$\text{Var}(T_k) = \frac{1}{t} \sum_{ij} e^2_{ij} . \tag{7}$$

Similarly we find:

$$\text{Cov}(T_k, T_{k'}) = E[(\sum_i \sum_j \delta^k_{ij} e_{ij})(\sum_i \sum_j \delta^{k'}_{ij} e_{ij})]$$

$$= E[\sum_i \sum_j \delta^k_{ij} \delta^{k'}_{ij} e^2_{ij} + \sum_i \sum_{j\neq j'} \delta^k_{ij} \delta^{k'}_{ij'} e_{ij} e_{ij'}$$

$$+ \sum_{i\neq i'} \sum_j \sum_{j'} \delta^k_{ij} \delta^{k'}_{i'j'} e_{ij} e_{i'j'}] .$$

The expectation of the first term is zero since $E(\delta^k_{ij} \delta^{k'}_{ij}) = 0$.

Also the expectation of the last term is zero since

$$E(\delta^k_{ij} \delta^{k'}_{i'j'}) = \frac{1}{t^2} , \quad \text{and} \quad \sum_j e_{ij} = 0 .$$

Hence:

$$\mathrm{Cov}(T_k, T_{k'}) = \frac{1}{t(t-1)} \sum_i \sum_{j' \neq j} e_{ij}\, e_{ij'}$$

$$= \frac{1}{t(t-1)} \sum_i [(\sum_j e_{ij})^2 - \sum_j e_{ij}^2] \quad \text{or:}$$

$$\mathrm{Cov}(T_k, T_{k'}) = \frac{-1}{t(t-1)} \sum_i \sum_j e_{ij}^2 \quad . \tag{8}$$

Using (7) and (8) we have:

$$V(T_k - T_{k'}) = \frac{2}{t-1} \sum_{ij} e_{ij}^2 \quad , \tag{9}$$

and similarly for the difference between two means:

$$V(\bar{T}_k - \bar{T}_{k'}) = \frac{2}{r^2(t-1)} \sum_{ij} e_{ij}^2 \quad . \tag{10}$$

1.2.2 The variance table:

Let us now find the expectations of the sums of squares included in the analysis of variance table:

Treatment

$$E(\text{treatment sum of squares}) = E[\frac{1}{r} \sum_k T_k^2 - \frac{(\sum_k T_k)^2}{rt}]$$

$$= E[\frac{1}{r} \sum_k T_k^2 - \mu^2 rt]$$

$$= \frac{1}{r} \sum_k E[r\mu + rt_k + \sum\sum_{ij} \delta_{ij}^k\, e_{ij}]^2 - \mu^2 rt$$

$$= tr\mu^2 + r\sum_k t_k^2 - \mu^2 rt + \frac{1}{r} \sum_k [2(r\mu + rt_k)\frac{1}{t} \sum_i \sum_j e_{ij}] + \frac{1}{r}\sum_k(\frac{1}{t} \sum_{ij} e_{ij}^2)$$

$$= r \sum_k t_k^2 + \frac{1}{r} \sum_i \sum_j e_{ij}^2 \quad . \tag{11}$$

Block

Denoting the total of the ith block by B_i we can write:

$$E(\text{Block sum of squares}) = E[\frac{1}{t} \sum_i B_i^2 - \mu^2 rt]$$

$$= \frac{1}{t} \sum_i E[t\mu + tb_i + \sum_k \sum_j \delta_{ij}^k \; e_{ij}]^2 - \mu^2 rt \quad .$$

But $\sum_k \sum_j \delta_{ij}^k \; e_{ij} = \sum_j e_{ij} \sum_k \delta_{ij}^k = \sum_j e_{ij} = 0$, since $\sum_k \delta_{ij}^k = 1$.

Hence:

$$E(SSB) = \frac{1}{t} \sum_i E[t\mu + tb_i]^2 - \mu^2 rt = t\sum b_i^2 \quad . \qquad (12)$$

Total

$$E(\text{Total sum of squares}) = E \sum_{ik} (y_{ij} - y_{..})^2$$

$$= \sum_{ik} E(b_i + t_k + \sum_j \delta_{ij}^k \; e_{ij})^2$$

$$= t\sum_i b_i^2 + r\sum_k t_k^2 + \sum_{ik} [\sum_j (\delta_{ij}^k)^2 \; e_{ij}^2 + \sum_{j \neq j'} \delta_{ij}^k \; \delta_{ij'}^k \; e_{ij} \; e_{ij'}] \quad .$$

Remembering that $\sum_i b_i = 0$, $\sum_j e_{ij} = 0$ we have:

$$E(\text{Total SS}) = t \sum_i b_i^2 + r \sum_k t_k^2 + \sum_{ij} e_{ij}^2 \quad . \qquad (13)$$

Error

$$E(\text{Error sum of squares}) = E(\text{Total SS}) - E(\text{Treatment SS}) - E(\text{Block SS})$$

$$= \frac{r-1}{r} \sum_{ij} e_{ij}^2 \quad . \qquad (14)$$

We can now construct the analysis of variance table as follows:

Table I

Source	Sum of squares	Expected sum of squares
Blocks	$t \sum_i (y_{i.} - y_{..})^2$	$t \sum_i b_i^2$
Treatments	$r \sum_k (y_{.k} - y_{..})^2$	$\frac{1}{r} \sum_{ij} e_{ij}^2 + r \sum t_k^2$
Error	$\sum_{ik} (y_{ik} - y_{i.} - y_{.k} + y_{..})^2$	$\frac{r-1}{r} \sum_{ij} e_{ij}^2$
Total	$\sum_{ik} (y_{ik} - y_{..})^2$	$\sum_{ij} e_{ij}^2 + r \sum t_k^2 + t \sum b_i^2$

If we divide the error sum of squares by (r-1)(t-1) we obtain a quantity whose expectation is $\frac{1}{r(t-1)} \sum_{ij} e_{ij}^2$. Denoting this by σ^2 we have by (10):

$$\mathrm{Var}(\hat{t_k - t_{k'}}) = \frac{2\sigma^2}{r} \quad . \tag{15}$$

Also:

$$E(\text{Treatment SS})/(t-1) = \frac{1}{r(t-1)} \sum_{ij} e_{ij}^2 + \frac{r}{t-1} \sum t_k^2 \quad . \tag{16}$$

If there are no treatment effects the expectation of the treatment mean square is equal to the expectation of the error mean square. This is the property of unbiasedness which is of paramount importance in any experimental design.

We see therefore that any comparison of the treatment effects $\Sigma \lambda_k t_k$ in which $\Sigma \lambda_k = 0$, is estimated unbiasedly by the same comparison of the observed treatment means. Moreover, for the variance,

$$V(\Sigma \lambda_k \hat{t}_k) = V(\Sigma \lambda_k \overline{T}_k)$$

$$= \Sigma \lambda_k^2 \operatorname{Var}(\overline{T}_k) + \sum_{k \neq k'}\sum \lambda_k \lambda_{k'} \operatorname{Cov}(\overline{T}_k, \overline{T}_{k'}) \quad .$$

Using (7) and (8), we therefore have:

$$V(\Sigma \lambda_k \hat{t}_k) = \Sigma \lambda_k^2 \frac{1}{r^2 t} \sum_{ij} e_{ij}^2 - \sum_{k \neq k'}\sum \lambda_k \lambda_{k'} \frac{1}{r^2 t(t-1)} \sum_{ij} e_{ij}^2$$

$$= [\frac{1}{r^2} \sum_{ij} e_{ij}^2] \{\frac{1}{t} \Sigma \lambda_k^2 - [(\Sigma \lambda_k)^2 - \Sigma \lambda_k^2] \frac{1}{t(t-1)}\}$$

$$= \frac{1}{r^2} \sum_{ij} e_{ij}^2 [\frac{1}{t} \Sigma \lambda_k^2 + \frac{1}{t(t-1)} \Sigma \lambda_k^2]$$

$$= \Sigma \lambda_k^2 \frac{\sigma^2}{r} \quad , \tag{17}$$

where $\sigma^2 = \frac{1}{r(t-1)} \sum_{ij} e_{ij}^2$.

It is to be noticed that we have not used any assumptions about the errors except that there are fixed deviations of plot values from block means, these deviations being attached at random to the treatment yields. We have used no assumptions of homogeneity of errors, where error refers to the quantity $\sum_j \delta_{ij}^k e_{ij}$, which is the error of the observed yield. Moreover, we have not assumed that $\sum_j e_{ij}^2$ is the same for all values of i.

1.2.3 Test of hypotheses:

Consider now the test of the hypothesis that the t_k's are all zero. For this purpose it is necessary to obtain the distribution of the treatment sum of squares and the distribution of the error sum of squares. From these we may obtain the distribution of the ratio:

(treatment mean square/error mean square) .

The derivation of these distributions over our finite population is a problem of considerable difficulty. First we should notice that we have in each block t! ways of assigning t treatments into t plots, one treatment per plot per block. Since we have r blocks we can get $(t!)^r$ plans for our randomized design. Hence the distribution of the treatment sum of squares, and the distribution of the error sum of squares are in fact the frequency functions of these quantities (in the absence of treatment effects) over the $(t!)^r$ possible plans. For example, suppose we have a simple experiment of 3 treatments and 3 blocks, we should have to calculate the sum of squares attributable to treatments, blocks, and errors in the 216 possible plans. This task is in general beyond practical consideration. Therefore it is expedient to resort to the infinite model and regard the e_{ij}'s as independent normal variates.

Referring to Table I, let:

$$S = r \sum_k (y_{.k} - y_{..})^2 + \sum_{ik} (y_{ik} - y_{i.} - y_{.k} + y_{..})^2 \tag{18}$$

$$T = r \sum_k (y_{.k} - y_{..})^2 \quad . \tag{19}$$

It should be noted that both the total sum of squares and block sum of squares retains the same value over all possible plans, whereas the treatment sum of squares and, consequently, the error sum of squares are variable from plan to plan. This is equivalent to saying that the quantity,

(Total SS) - (Block SS) = (Treatment SS)+(Error SS)

has the same value in each plan. Hence, as far as the variation over the $(t!)^r$ possible plans is concerned, the quantity S is constant. On the contrary, the quantity T is a random variable. Now,

$$E(S) = E\{(\text{Treatment SS}) + (\text{Error SS})\}$$

$$= E(\text{Treatment SS}) + E(\text{Error SS})$$

$$= \frac{1}{r} \sum_{ij} e_{ij}^2 + \frac{r-1}{r} \sum_{ij} e_{ij}^2 \quad \text{(under the null hypothesis)}$$

$$= \sum_{ij} e_{ij}^2 = S \quad . \tag{20}$$

Under the normality assumption, the ratio:

$$\frac{T}{t-1} \Big/ \frac{S - T}{(r-1)(t-1)} \tag{21}$$

would follow the F-distribution, and

$$x = \frac{T}{S} \tag{22}$$

would follow the β-distribution by the well-known relation,

$$F = \frac{f_2 B}{f_1(1-B)} \quad , \tag{23}$$

where f_1 and f_2 are the degrees of freedom.

The probability density of the variate x is given by:

$$f(x) = \frac{\Gamma\frac{1}{2}(m+n)}{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2}n)} x^{\frac{1}{2}m-1} (1-x)^{\frac{1}{2}n-1} \quad , \tag{24}$$

where m(=t-1) and n(=(r-1)(t-1)) are the degrees of freedom for T and S-T, respectively.

But from (24):

$$E(x) = \frac{m}{n+m} = \frac{1}{r} \tag{25}$$

$$\mathrm{Var}(x) = \frac{2mn}{(m+n)^2(m+n+2)}$$

$$= \frac{2(r-1)}{r^2(t-1)(rt-r+2)} \quad . \tag{26}$$

In contrast, for the finite model we have,

$$E(T) = \frac{S}{r} \quad , \tag{27}$$

under the null hypothesis that there are no treatment effects.

Moreover,

$$\mathrm{Var}(T) = \frac{2}{r^2(t-1)} (S^2-K) \quad , \tag{28}$$

where

$$K = \sum_i (\sum_j e_{ij}^2)^2 \quad .$$

Now if the errors associated with the blocks are homogeneous,

$$\sum_j e_{ij}^2 = \frac{S}{r} \qquad \text{(independent of i)} ,$$

and thus $k = \frac{S^2}{r}$ giving:

$$\mathrm{Var}(T) = 2(r-1)S^2/(t-1)r^3 \quad ,$$

and consequently

$$\mathrm{Var}(\frac{T}{S}) = \frac{1}{S^2} \mathrm{Var}(T) = \frac{2(r-1)}{(t-1)r^3} \quad . \tag{29}$$

Comparing the normal theory result for Var(x) given in (26) with (29) shows remarkable agreement, assuming $\frac{2}{r(t-1)}$ is small. We may therefore conclude that if the error variance is the same for all blocks then the distribution of $\frac{T}{S}$ is nearly that of the β-distribution in (24). Hence, in view of (23) the distribution of the criterion ratio (21) is very nearly the F-distribution. This result is due to Welch and Pitman (1937).

It is of interest to notice that heterogeneity of errors does not effect the mean of the distribution of $\frac{T}{S}$ but does effect the variance. For denoting $\sum_j e_{ij}^2$ by S_i and $\Sigma S_i/r$ by $\overline{S}$, we have

$$\begin{aligned} Var(\frac{T}{S}) &= 2[S^2 - \Sigma S_i^2]/[(t-1)r^2 S^2] \\ &= \frac{2(r-1)}{(t-1)r^3} [1 - \frac{\Sigma(S_i-\overline{S})^2}{r(r-1)\overline{S}^2}] \qquad (30) \\ &< \frac{2(r-1)}{(t-1)r^3} . \end{aligned}$$

Thus the heterogeneity of errors actually reduces the variance.

1.2.4 Remarks: The above example throws some light on the role played by the assumptions of "additivity" and "homogeneity". The further assumption of "normality" makes it possible to carry out a fairly accurate analysis of randomized designs of this type by an appeal to the usual normal theory and F-test procedures.

1.3 Effect of non-additivity: According to Cochran (1947), the principal effect of the failure of the assumption of additivity is a loss of information. He considers the usual error variance σ_e^2 as inflated by the component σ_{na}^2 due to non-additivity, and regards the ratio $\frac{\sigma_{na}^2}{\sigma_e^2+\sigma_{na}^2}$ as a measure of the loss of information due to non-additivity. Thus we can say that when treatment and replicate effects are small the loss of information will be trivial unless σ_e is very small. But when either the treatment or the replicate effect is substantial the loss of information may be important, and when both effects are large the loss of information may be considerable. Cochran states that the loss of information should be negligible when treatment and replicate effects do not exceed 20 percent of the total variation.

1.4 Effect of heterogeneity: In the example considered in section (1.2), let us suppose that we have an experiment in 4 blocks and that the variance of one of the blocks is 3 times that of each of the other blocks (assumed equal). Then using (30) we find that the variance of T/S is

$$\frac{2(r-1)}{(t-1)r^3}\left(1 - \frac{1}{9}\right) = \frac{2(r-1)}{(t-1)r^3} \cdot \frac{8}{9} \quad .$$

If in addition we suppose that there are 6 treatments, the parameters m and n of the approximating Beta-distribution given in (24) are given by the equations

$$\frac{m}{m+n} = \frac{1}{r} = \frac{1}{4}$$

$$\frac{2mn}{(m+n)^3} = \frac{2(r-1)}{(t-1)r^3} \cdot \frac{8}{9} = \frac{1}{60} \quad ,$$

for which the solution is $m = \frac{360}{64} = 6$ approximately, $n = 3m = \frac{1080}{64} = 17$ approximately. In order to obtain an approximation to the level of significance using the randomization test, we should compare the significance levels of the F-distribution using 6 and 17 degrees of freedom as against using 5 and 15 degrees of freedom. Hence the ordinary use of the F-distribution, when heterogeneous errors are present, underestimates the level of significance of the randomization test.

Cochran (1947) states that, if in the ordinary analysis of variance the error variances are heterogeneous, then there will be a loss of efficiency in the estimates of treatment effects. Moreover, there will be a loss of sensitivity in the tests of significance.

1.5 <u>Effects of non-normality</u>: The effect of non-normality is not generally considered to be serious for moderate departures from normality. But, any large departures from normality in the region of the outlying observations are likely to affect the validity of significance tests. Moreover non-normality is likely to be accompanied by a loss of efficiency in the estimation of treatment effects and a corresponding loss of power in the F- and t-tests.

1.6 <u>Properties of ideal transformation</u>:

Bartlett defines the ideal transformation as one having the following four properties:

a. The variance of the transformed variate should be unaffected by changes in the mean level.

b. The transformed variate should be normally distributed.

c. The transformed scale should be one for which an arithmetic average is an efficient estimate of the true mean level for any particular group of measurements.

d. The transformed scale should be one for which real effects are linear and additive.

Although these conditions are to some extent related [for example, (a) and (b) and (d) together imply (c)], we obviously cannot necessarily expect to arrange for conditions (b), (c), and (d) to be satisfied if our scale has already been fixed by condition (a). Fortunately the transformation of scale to meet condition (a) often has the effect of improving the closeness of the distribution to normality. A correlation of variance with the mean, on the original scale, often implies excessive skewness which tends to be eliminated after the transformation.

Condition (c) is required because the estimates which arise in the analysis of variance are of the simple arithmetic average type and we want to know whether such estimates are efficient. In view of the asymptotic nature of the transformation problem, we can define the efficiency to be $100r^2$ where (r) is the correlation between the transformed and the original scales.

The contention is sometimes made that the original scale is the more relevant one for taking sums and averages, and more understandable. While this argument has some force and is a warning against making transformations without good reason, it loses strength when we remember that if the variability in the data varies with the mean level for different blocks or groups, an unweighted average of the observed treatment responses is not necessarily the best estimate of the true treatment response, and the average on the transformed scale will often be the better estimate when reconverted to the original scale.

Condition (d) is required because it is more effective and simpler to be working on a scale for which treatment or other effects are linear and additive. But it is not always possible to choose a scale to cover condition (a) and yet be most reasonable for (d), though it may happen that a choice of scale for (a) improves the scale to some extent as far as (d) is concerned.

Practical considerations of some transformations

1.7 The angular transformation:

1.7.1 Introduction: If we draw independently a random sample of n items from an infinite population in which a proportion P of the items possess some definite attribute A, then the number of items (X) possessing the attribute A in our sample is distributed as a Binomial with parameters (n,P). Let p denote the observed proportion of items possessing the attribute A, then

$$E(p) = P \tag{31}$$

and

$$Var(p) = \frac{P(1-P)}{n} \quad . \tag{32}$$

The expression (32) is an explicit relation between the variance and the mean of the observed proportion p, which makes data consisting of proportions or percentages not amenable to the analysis of variance.

Recognizing this fact, R. A. Fisher (1922) in a paper on theoretical genetics introduced the transformation

$\cos \varphi = 1-2p$, i.e., $p = \sin^2 \frac{\varphi}{2}$; where φ varies from zero to π as p varies from zero to one. Not until 1936 was the general applicability of this transformation discussed in print, when Bartlett on a suggestion from Yates considered the use of a related transformation ($\text{Si}^2\,\theta = p$, where θ varies from zero to $\frac{\pi}{2}$) as a means of rendering proportions amenable to the analysis of variance.

1.7.2 <u>Development of the transformation</u>:

Let the transformation be f(p), then

$$E\, f(p) = E\, f(P+h) \qquad\qquad (h = p-P)$$

$$= E[f(P) + h\, f'(P) + R(h)]$$

$$= f(P) \quad \text{(approximately)} \quad ,$$

assuming we can ignore E R(h). Then

$$\text{Var } f(p) = E[f(p) - f(P)]^2 \qquad \text{(approximately)}$$

$$= [f'(P)]^2 \cdot \text{Var}(p) \quad .$$

Hence, if Var f(p) is to be independent of P (or nearly so), then by (2):

$$f'(P) = \frac{c}{\sqrt{\text{Var}(p)}} = \frac{c}{\sqrt{\frac{P(1-P)}{n}}}$$

$$= \frac{1}{\sqrt{P(1-P)}} \qquad \text{(taking the arbitrary constant } c \text{ as equal } \frac{1}{\sqrt{n}}\text{)} \quad .$$

It follows that

$$f(p) = \int_0^p \frac{dP}{\sqrt{P(1-P)}}$$

$$= 2 \text{ arc sin } \sqrt{p} \quad . \tag{33}$$

1.7.3 Study of the bias and the variance:

1.7.3.0 Let us take $2 \sin^{-1} \sqrt{p} = \varphi$ and $2 \sin^{-1} \sqrt{P} = \phi$, where p and P are the observed and theoretical proportions respectively. We wish to determine $E(\varphi-\phi)$ and $E(\varphi-\phi)^2$.

Let h be the deviation p - P. By Taylor's theorem we can write:

$$\varphi = 2f(p) = 2f(P+h) = 2[f(P)+h\ f^{(1)}(P) + \frac{h^2}{2!} f^{(2)}(P) + \frac{h^3}{3!} f^{(3)}(P) + \frac{h^4}{4!} f^{(4)}(P)+\ldots]$$

and

$$\varphi-\phi = 2[h\ f^{(1)}(P) + \frac{h^2}{2!} f^{(2)}(P) + \frac{h^3}{3!} f^{(3)}(P) + \frac{h^4}{4!} f^{(4)}(P)+\ldots] \quad . \tag{34}$$

Now

$$f^{(1)}(P) = [\frac{d}{dp} (\sin^{-1} \sqrt{p}\ )]_{p=P}$$

$$= \frac{1}{2\sqrt{P-P^2}} \tag{35}$$

$$f^{(2)}(P) = -\frac{(1-2P)}{4(P-P^2)^{3/2}} \tag{36}$$

$$f^{(3)}(P) = \frac{8P^2-8P+3}{8(P-P^2)^{5/2}} \tag{37}$$

$$f^{(4)}(P) = \frac{3(1-2P)(-8P^2+8P-5)}{16(P-P^2)^{7/2}} \quad . \tag{38}$$

Thus

$$E(\varphi-\phi) = \sum_{r=1}^{\infty} 2\left(\frac{d^r}{dp} \sin^{-1} \sqrt{p}\right)E(h^r) \quad . \tag{39}$$

But

$$\begin{aligned} E(h) &= 0 \\ E(h^2) &= \frac{P(1-P)}{n} \\ E(h^3) &= \frac{P(1-P)(1-2P)}{n^2} \\ E(h^4) &= \frac{3P^2(1-P)^2}{n^2} + \frac{P(1-P)(1-6P+6P^2)}{n^3} \quad . \end{aligned} \tag{40}$$

Substituting (35), (36), (37), (38), and (40) into (39) we get:

$$E(\varphi-\phi) = -\frac{1-2P}{4n(PQ)^{1/2}} - \frac{(1-2P)(21-8P+8P^2)}{192(PQ)^{3/2}\, n^2} + 0(n^{-2}) \quad . \tag{41}$$

1.7.3.1 If n is sufficiently large ($>$ 30) then the terms involving $\frac{1}{n}$ to higher power than the first can be neglected provided $0 < P_1 < P < P_2 < 1$, i.e., P is somewhat far from the end points zero and one. There will still remain a bias in φ which equals $[-\frac{1-2P}{4n(P-P^2)^{\frac{1}{2}}}]$ asymptotically. This quantity however is divided by n and becomes negligible when n is sufficiently large. It hould also be noted that when $P > .5$, $E(\varphi-\phi) > 0$, and when $P < .5$, $E(\varphi-\phi) < 0$. For $P = .5$, $E(\varphi-\phi) = 0$.

1.7.3.2 In order to prove that $E(\varphi) = \phi$ when $P = .5$, Dr. Shenton suggested the following nice proof:

Let us compute $E \sin^{-1} \sqrt{p}$ for $P = .5$; we get:

$$E \sin^{-1} \sqrt{p} = Pr(0) \sin^{-1} \sqrt{0} + Pr(1) \sin^{-1} \sqrt{\frac{1}{n}} \dots + Pr(n) \sin^{-1} \sqrt{\frac{n}{n}}$$

$$= \frac{1}{2^n}[n \sin^{-1}\sqrt{\frac{1}{n}} + \frac{n(n-1)}{2}\sin^{-1}\sqrt{\frac{2}{n}} \dots + \frac{\pi}{2}] \quad .$$

In view of the symmetricity of Binomial coefficients and of the facts that $\sin^{-1}\sqrt{\frac{1}{n}}+\sin^{-1}\sqrt{\frac{n-1}{n}} = \frac{\pi}{2}$, we get for both $n = (2N+1)$ (odd) and $n = 2N$ (even) the following results:

$$E(\sin^{-1}\sqrt{p}) = \frac{1}{2^n}[\frac{\pi}{2} + \frac{n}{1}\frac{\pi}{2} + \binom{n}{2}\frac{\pi}{2} \dots + \binom{2N+1}{N}\frac{\pi}{2}]$$

$$= \frac{1}{2^n} \cdot \frac{\pi}{2}[1 + \binom{2N+1}{1} + \binom{2N+1}{2} \dots + \binom{2N+1}{N}]$$

$$= \frac{1}{2^n} \cdot \frac{\pi}{4}(1+1)^{2N+1} = \frac{\pi}{4} = \sin^{-1}\sqrt{P} \quad \text{for } P=.5 \quad .$$

For $n = 2N$ we have:

$$E(\sin^{-1}\sqrt{p}) = \frac{1}{2^n}[\frac{\pi}{2} + \binom{2N}{1}\frac{\pi}{2} + \binom{2N}{2}\frac{\pi}{2} \ldots + \binom{2N}{N-1}\frac{\pi}{2} + \binom{2N}{N}\frac{\pi}{4}]$$

$$= \frac{1}{2^n} \cdot \frac{\pi}{2}[1 + \binom{2N}{1} + \binom{2N}{2} \ldots + \binom{2N}{N}\frac{1}{2}] \quad .$$

Between brackets we have the first half of the complete Binomial expansion $(1+1)^n$. Hence:

$$E(\sin^{-1}\sqrt{p}) = \frac{1}{2^n} \cdot \frac{\pi}{2} \cdot \frac{1}{2}(1+1)^n$$

$$= \frac{\pi}{4} = \sin^{-1}\sqrt{P} \quad .$$

1.7.3.3 Turning our attention to the variance, we have

$$E(\varphi-\phi)^2 = 4E\{h^2[f^{(1)}(P)]^2 + h^3 f^{(1)}(P) f^{(2)}(P)$$

$$+ h^4[\frac{(f^{(2)}(P))^2}{4} + \frac{f^{(1)}(P) f^{(3)}(P)}{3}]\}$$

$$= 4\{\frac{1}{4PQ}\frac{PQ}{n} - \frac{(1-2P)^2}{2n^2PQ} + \frac{12P^2Q^2}{n^2}[\frac{3(1-2P)^2+4(8P^2-8P+3)}{192(PQ)^3}]\}$$

$$= \frac{1}{n} + \frac{12P^2-12P+7}{16n^2PQ} + 0(n^{-2}) \quad . \qquad (42)$$

$E(\varphi-\phi)^2$ is approximately the variance of φ and is equal to $\frac{1}{n}$, ignoring terms of order n^{-2} and higher.

In most field experiments the percentages are based upon relatively large numbers of 100 or more individuals, so that usually the transformation would accomplish its purpose, avoiding the end points of the range of p. The effect of the end points on the variance of φ has been discussed by M. S. Bartlett whose work will be referred to in a later section.

1.7.4 Types of percentage data:

The type of data for which the angular transformation is suitable is that having a Binomial nature. This means that the data is discrete in the first place and based upon a determinate number of trials (n) say. Data of this kind is quite often expressed as percentages. Examples of this kind of data are: Germination percentages given by number of seeds germinated/total seeds, or disease percentages given by number of plants diseased/total plants, etc.

Data for which the angular transformation should not be applied consists of continuous data arising from an experimental study in which results are given as percentages. Thus, each variate may be divided by an arbitrary constant value yielding percentages of some standard or average. Clearly such a procedure merely transforms the unit of measurement. For example yield data might be expressed as a percentage of the control reading instead of actual yield.

Again concentrations are often expressed as percentages since a comparative assessment is the object of the experimental design. This type of percentage is very common; for example, (a) seed purity given by weight of pure seed/total weight of seed, (b) protein content given by weight of protein/total weight, and (c) sugar content given by weight of sugar/weight of root. Such concentrations should not as a rule be subjected to any transformation to stabilize the variance. However the technique applied to problems involving percentage data must be carefully considered in deciding which transformation to use.

1.8 Practical investigations involving square root and angular transformations:

1.8.1 Development of the square root transformation: Suppose we have a variate x whose variance is a function of its mean, and we wish to find a transformation that stabilizes the variance. Let the transformation be $f(x)$ and assume $E(x) = \mu$. Then

$$f(x) = f(\mu) + \frac{x-\mu}{1} f'(\mu) + R \quad ,$$

and

$$E\, f(x) = f(\mu) \quad , \tag{43}$$

assuming we can ignore E R(x). Then

$$\text{Var } f(x) = [f'(\mu)]^2 \, E(x-\mu)^2$$

$$= \text{Var}(x)[f'(\mu)]^2 \quad . \tag{44}$$

Hence if Var f(x) is to be independent of x (or nearly so), then

$$\text{Var } f(x) = \text{Var}(x)[f'(\mu)]^2 = c^2 \quad .$$

Let Var(x) = g(μ), then

$$f'(\mu) = \frac{c}{\sqrt{g(\mu)}}$$

and

$$f = \int \frac{c\,d\mu}{\sqrt{g(\mu)}} \quad . \tag{45}$$

If $g(\mu) = \mu$,

$$f(x) = \int \frac{c\,dx}{\sqrt{x}} = \sqrt{x} \tag{46}$$

provided $c = \frac{1}{2}$.

This transformation is the square root transformation. The variance is $c^2 = \frac{1}{4}$ provided the mean is large enough to enable us to ignore the remainder R mentioned above. In general if $\text{Var}(x) = \lambda\mu$, the square root transformation will give a variance $= \frac{1}{4}\lambda$.

1.8.2 Empirical study of variance stabilization in the Poisson case:

From above we see that the Poisson distribution suggests itself as a distribution for which the square root transformation may prove useful, and replicated experiments where the results are Poisson variates may often be analyzed in this way. Since the approximation above depends on the magnitude of the mean, say m, it is interesting to see how far the variance of $\sqrt{x}$ for a Poisson variate x is constant when the mean m becomes small. Bartlett (1936) investigated this point and calculated for the sake of comparison the variance of $\sqrt{x}$ for a continuous distribution for which also the mean equals the variance, namely the Type III distribution

$$P \propto x^{m-1} e^{-x} dx \quad . \tag{47}$$

In view of the discontinuous nature of the Poisson distribution, the variance of $\sqrt{x+\frac{1}{2}}$ was further found for the Poisson distribution. The results of Bartlett's investigations are given in Table II and Figure I.

The variance for the continuous curve approaches its limit (.25) surprisingly quick. That for $\sqrt{x}$ is reasonably convergent, but shows a peak round about m=1, which disappears when $\sqrt{x+\frac{1}{2}}$ is used.

Table II. Comparison of Variances with Increasing m.

mean m	Poisson, $\sqrt{x}$	Poisson, $\sqrt{x+\frac{1}{2}}$	Continuous, $\sqrt{x}$
.0	.000	.000	.000
.5	.310	.102	.182
1.0	.402	.160	.215
2.0	.390	.214	.233
3.0	.340	.232	.239
4.0	.306	.240	.242
6.0	.276	.245	.245
9.0	.263	.247	.247
12.0	.259	.248	.249
15.0	.256	.248	.250

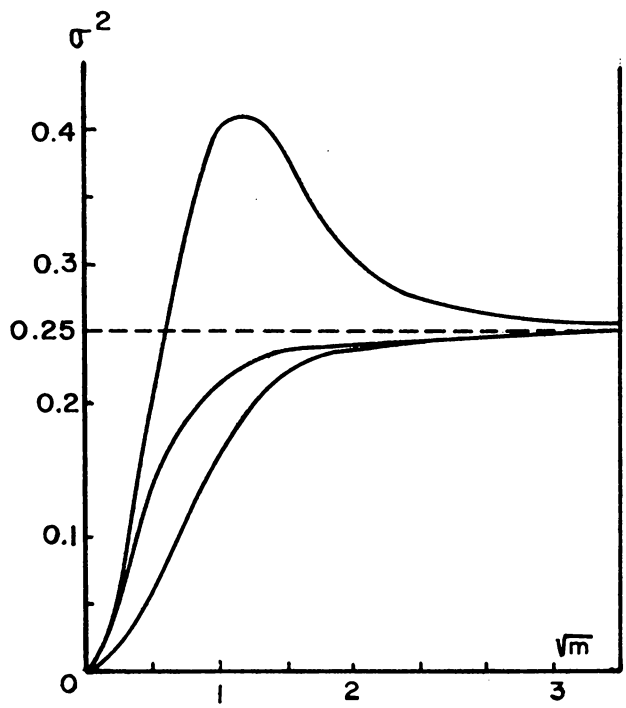

Fig. I. The three curves (the largest variance first) are for: 1. Poisson, $\sqrt{x}$,
2. Type III, $\sqrt{x}$,
3. Poisson, $\sqrt{x+\frac{1}{2}}$.

Thus above a mean of 10, say, $\sqrt{x}$ may be considered; from 10 to about 2 or 3, $\sqrt{x+\frac{1}{2}}$ is preferable. Below a mean of about 2 or 3 the discontinuous nature of the Poisson distribution has, of course, become so violent that any variate is of little use quite apart from questions of variance, unless a large number of replications is available.

1.8.3 Efficiency of square root transformation: Bartlett states that it is hardly possible to consider very fully the question of efficiency without carefully specifying the nature of the experiment to be analyzed. The efficiency depends, for example, on how we choose to define our treatment effects. As a rough guide, however, we may note that the value of r^2, where r is the correlation of $\sqrt{x}$ or $\sqrt{x+\frac{1}{2}}$ with x for a Poisson distribution, is high. Thus the percent efficiency $100r^2$ of the total $(\Sigma \sqrt{x})$ in large samples for estimating m, in comparison with the sufficient statistic Σx, has a minimum of about 88 percent at about $m = 2$; that of $\sqrt{x+\frac{1}{2}}$ falls to $96\frac{1}{2}$ percent at the same value of m. In so far as these figures differ, they favour $\sqrt{x+\frac{1}{2}}$. The reason for computing the efficiency is that we want to have some confidence that the treatment means given in terms of the new variate do adequately summarize our data, and this they could not do if their efficiency for estimating the true parameter means m on the original scale was low.

The transformation will probably make our variate nearer normal than before. However, this point will be discussed in greater detail in the sequel.

If m is very small but a fairly large number of replications is available, it might sometimes be advisable to add, say, pairs of replications before square root are taken.

1.8.4 Binomial type of data - stabilization of variance in small samples:

We have found that for a Binomial type of data the angular transformation is effective, but no work has been done on the question of how large a sample size n must be used to ensure a reliable transformation. Bartlett (1936) investigated the effectiveness of the inverse sine transformation in small samples. He regards n = 10 as a lower limit for n because it is hardly likely that any experiment would be carried out with $n < 10$. The variance of $y = \frac{2}{\pi} \sin^{-1}\sqrt{\frac{t}{n}}$ for n = 10 was examined together with the variance of the corresponding continuous Type I distribution,

$$P \propto x^{\alpha-1} (1-x)^{8-\alpha} dx \quad . \tag{48}$$

The mean in (48) is $\frac{\alpha}{9}$ and the variance is $\frac{\frac{\alpha}{9}(1-\frac{\alpha}{9})}{10} = \frac{m(1-m)}{10}$ (m is the mean) so that we have exactly the same relation between the mean and the variance as in a Binomial with n = 10. By analogy with $\sqrt{x}$ the variance of $z = \frac{2}{\pi} \sin^{-1}\sqrt{\frac{t \pm \frac{1}{2}}{10}}$ was also found, t being the Binomial variate, and ½ being added or subtracted according as $t \lessgtr 5$. The binomial variances were computed directly. For the continuous case we have:

$$E(\sin^{-1}\sqrt{x}) = \int_0^1 \sin^{-1}\sqrt{x}\ Pdx$$

$$= [\sin^{-1}\sqrt{x} \int pdx]_0^1 - \int_0^1 \tfrac{1}{2}x^{-\frac{1}{2}}(1-x)^{-\frac{1}{2}}[\int pdx]\,dx$$

and may be evaluated for α and β $(=9-\alpha)$ integers.

Further we have:

$$E(\sin^{-1}\sqrt{x})^2 = \int_0^1 (\sin^{-1}\sqrt{x})^2\ Pdx$$

$$= [(\sin^{-1}\sqrt{x})^2 \int Pdx]_0^1 - \int_0^1 \sin^{-1}\sqrt{x}$$

$$\cdot\ x^{-\frac{1}{2}}(1-x)^{-\frac{1}{2}}[\int Pdx]\,dx \quad .$$

Expanding $f = \sin^{-1}\sqrt{x}$ in a Maclaurin series in $\sqrt{x}$, we get:

$$f = x^{1/2} + \frac{x^{3/2}}{3!} + \frac{3^2 x^{5/2}}{5!} + \frac{3^2 5^2 x^{7/2}}{7!} + \ldots$$

which can be used to evaluate the second integral. The results are given in Table III for $\alpha = 0$ to 4 and in Fig. II.

The curves for the Binomial depart considerably from that for the continuous distribution. That for y has a discrepancy at the two ends analogous to that for $\sqrt{x}$ in the case of a Poisson variate. The curve for z shows the anticipated drop in the center due to the use of $\pm\frac{1}{2}$. For larger n, the curve z will become more stable. The large sample efficiencies of y and z, like those of $\sqrt{x_1}$ and $\sqrt{x+\frac{1}{2}}$ remain high, that of y having a minimum of about 91 percent in the neighbourhood of $m = \frac{1}{9}$. Moreover, the efficiency for z is about 96 percent in the neighbourhood of the middle of the range. For the two variates y and z, z would naturally be used if the observed limit of the range was near the end points, otherwise y would be preferred. Fig. II shows how, when n = 10, the variance of y depends very markedly on the true proportion P.

1.8.5 <u>Further consideration of the effectiveness of the angular transformation</u>: We have mentioned before that when P is near to 0 or to 1, the transformation is not useful for small samples. Bartlett (1936) remarks that a

Table III. Comparison of Variances

mean m	Binomial y	Binomial z	Continuous y
0	.0	.0	.0
1/9	.0170	.0076	.0100
2/9	.0149	.0091	.0108
3/9	.0126	.0081	.0111
4/9	.0117	.0067	.0112

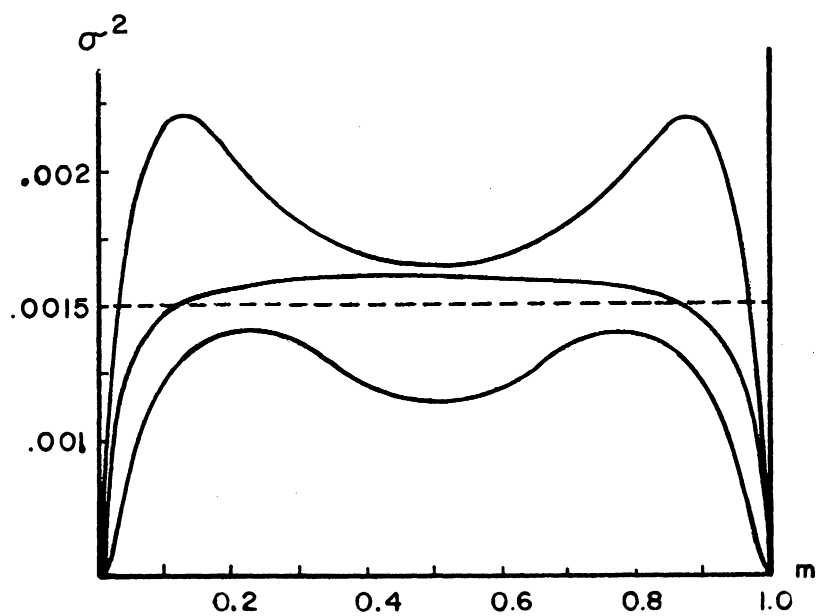

Fig. II. Change in variance with m.

1. Binomial (n = 10), y.
2. Type I ($\alpha + \beta = 9$), y.
3. Binomial (n = 10), z.

good correction is to replace $P = 0$ by $P = \frac{1}{4n}$, and $P = 1$ by $P = 1 - \frac{1}{4n}$. Tables IV, V, VI and the corresponding Figures III, IV, and V were given by Eisenhart (1947) showing a comparison between $\varphi(p) = 2 \arcsin\sqrt{p} = \arccos(1-2p)$, $(0 \leqq p \leqq 1,\ 0 \leqq \varphi \leqq \pi)$ and the adjusted $\varphi(p)$, namely $\varphi_B(p) = \varphi(p)$ for $(0 < p < 1)$, but $\varphi_B(0 = \varphi(\frac{1}{4n})$, $\varphi_B(1) = \pi - \varphi_B(0)$. The sample sizes were $n = 10, 20, 30$, respectively. The curve for Var(p) is drawn for comparison in each figure.

If the inverse sine transformation completely stabilized the variance in small samples, then $nv(\varphi)$ would be unity in all cases, or at least would be constant for a given value of n. This is not the case, as is apparent from the tables IV to VI. Generally the following conclusions may be drawn:

1. For $n \geqq 10$ and $.05 \leqq P \leqq .95$, $nv(\varphi)$ and $nv(\varphi_B)$ are less variable than $4nv(P)$.

2. For $n \geqq 20$ and $.05 \leqq P \leqq .95$, $nv(\varphi_B)$ is less variable than $nv(\varphi)$.

3. For $n \geqq 10$ and $.075 \leqq P \leqq .925$, $nv(\varphi_B)$ is less variable than $nv(\varphi)$.

It is quite apparent from the foregoing that the inverse sign transformation, especially with Bartlett's correction, is advantageous from the viewpoint of variance stabilization,

Table IV. Mean and variances of various functions of a sample proportion for selected values of the population proportion when the sample size = 10.

P=E(p)	φ(P)	E(φ)	10 v(φ)	$E(\varphi_B)$	$10\ v(\varphi_B)$
.0500	.4510	.2856	1.2917	0.4757	0.4481
.1000	.6435	.5134	1.6567	0.6241	0.7489
.1500	.7954	.7006	1.6529	0.7632	0.9363
.2000	.9273	.8598	1.5332	0.8939	1.0436
.3000	1.1593	1.1260	1.2968	1.1350	1.1224
.5000	1.5708	1.5708	1.1433	1.5708	1.1258
.7000	1.9823	2.0156	1.2968	2.0066	1.1224
.8000	2.2143	2.2818	1.5332	2.2477	1.0436
.8500	2.3462	2.4410	1.6529	2.3784	0.9363
.9000	2.4981	2.6282	1.6567	2.5175	0.7489
.9500	2.6906	2.8560	1.2917	2.6659	0.4481

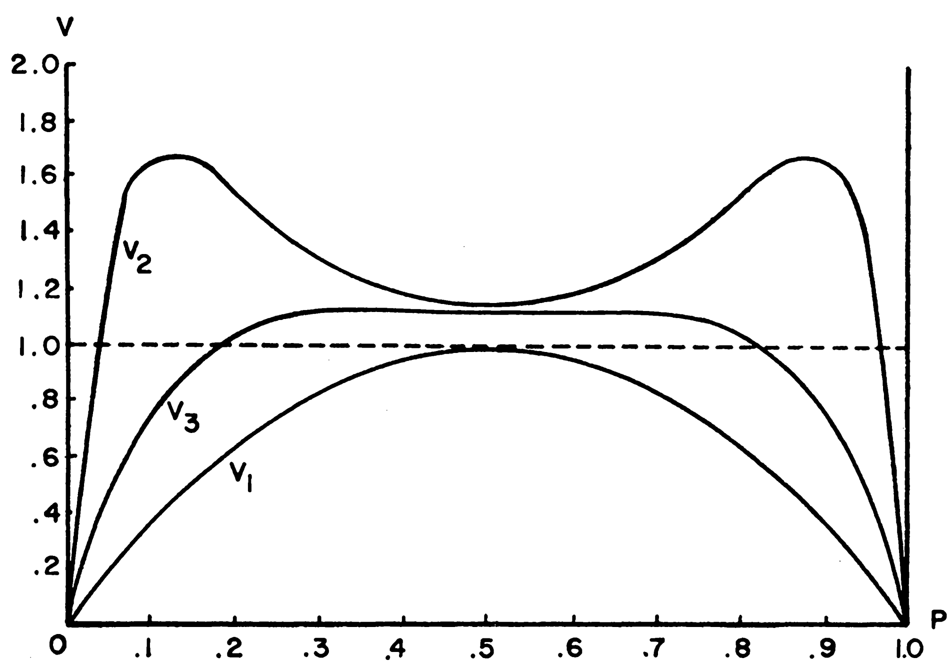

Fig. III. Dependence of the variances of 3 functions of the sample proportion p on the population proportion P when the sample size = 10.

$v_1=4P(1-P)=4 \times 10\ v(p)$, $v_2=10\ v(\varphi)$, $v_3= 10\ v(\varphi_B)$

Table V. Means and variances when the sample size is 20.

$P=E(p)$	$\varphi(P)$	$E(\varphi)$	$20\ v(\varphi)$	$E(\varphi_B)$	$20\ v(\varphi_B)$
.0500	.4510	.3541	1.6315	.4344	0.7247
.1000	.6435	.5866	1.5458	.6139	1.0138
.1500	.7954	.7602	1.3240	.7688	1.0974
.2000	.9273	.9035	1.1810	.9061	1.1051
.3000	1.1593	1.1469	1.0858	1.1471	1.0784
.5000	1.5708	1.5708	1.0562	1.5708	1.0561
.7000	1.9823	1.9947	1.0858	1.9945	1.0784
.8000	2.2143	2.2381	1.1810	2.2355	1.1051
.8500	2.3462	2.3814	1.3240	2.3728	1.0974
.9000	2.4981	2.5550	1.5458	2.5277	1.0138
.9500	2.6906	2.7875	1.6315	2.7072	0.7247

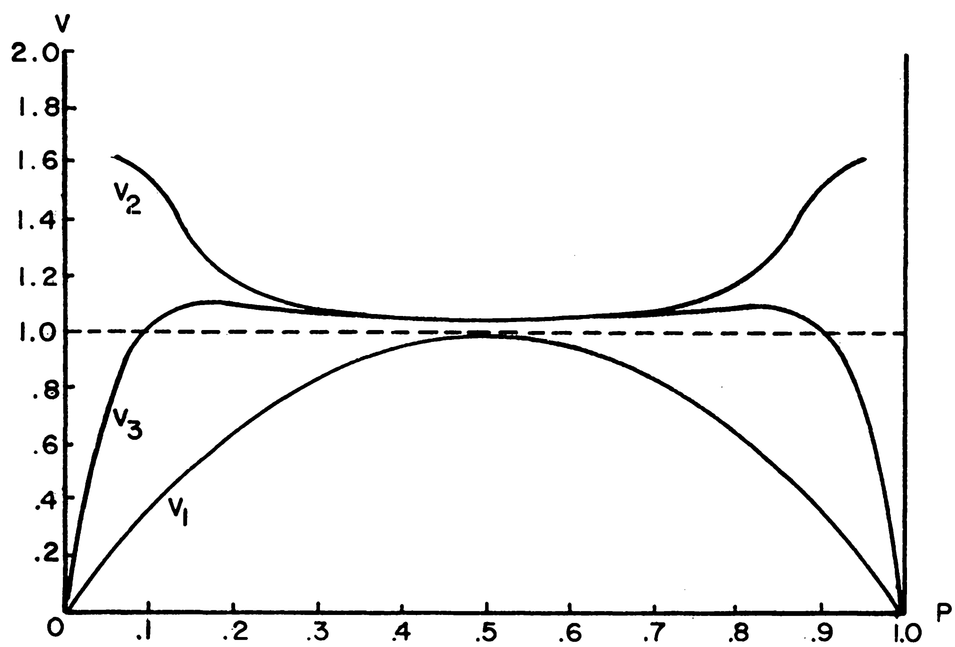

Fig. IV. Dependence of the variances on P.

$v_1 = 4P(1-P) = 4 \times 20\ v(P)$

$v_2 = 20\ v(\varphi)$

$v_3 = 20\ v(\varphi_B)$

Table VI. Mean and variances when the sample size is 30.

P=E(p)	φ(P)	E(φ)	30 v(φ)	$E(\varphi_B)$	30 $v(\varphi_B)$
.0500	.4510	.3885	1.6443	.4278	.8985
.1000	.6435	.6119	1.3359	.6197	1.0921
.1500	.7954	.7757	1.1560	.7771	1.0987
.2000	.9273	.9134	1.0885	.9163	1.0773
.3000	1.1593	1.1516	1.0484	1.1516	1.0481
.5000	1.5708	1.5708	1.0358	1.5708	1.0358
.7000	1.9823	1.9900	1.0484	1.9900	1.0481
.8000	2.2143	2.2282	1.0885	2.2280	1.0773
.8500	2.3462	2.3659	1.1560	2.3645	1.0987
.9000	2.4982	2.5297	1.3359	2.5219	1.0921
.9500	2.6806	2.7531	1.6443	2.7138	.8985

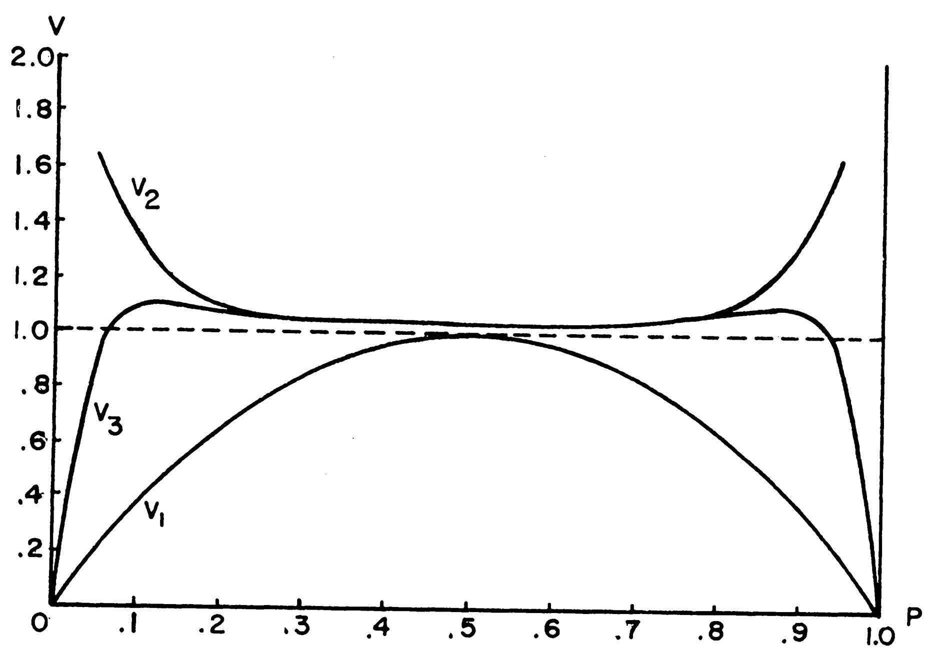

Fig. V. Dependence of the variances on P.

$v_1 = 4P(1-P) = 4 \times 30\ v(p)$

$v_2 = 30\ v(\varphi)$

$v_3 = 30\ v(\varphi_B)$

its effectiveness in this regard increasing rapidly as the sample size n is increased. A comparison of the values of $\varphi(P)$, $E(\varphi)$, and $E(\varphi_B)$ shown in the above tables reveals that $E(\varphi)$ and $E(\varphi_B)$ generally are less than $\varphi(P)$ when $P < \frac{1}{2}$, greater than $\varphi(P)$ when $P > \frac{1}{2}$. The three are equal when $P = .5$, in agreement with what we have found theoretically in section 1.7.3. Furthermore, $E(\varphi_B)$ is always closer to $\varphi(P)$. In fact when $.05 \leqq P \leqq .95$, the bias $E(\varphi_B) - \varphi(P)$ is less than 15 percent of the standard deviation of $\varphi(B)$ for $n \geqq 10$, and thus is of negligible magnitude for most practical purposes. Thus Bartlett's correction tends to reduce the biases to negligible quantities.

In conclusion, it must be remembered that while in large samples the variance of a transformed proportion is independent of the true proportion for most practical purposes, the variance of the transformed value φ is still proportional to $\frac{1}{n}$, where n is the number of observations upon which the observed proportion is based. Consequently, if $p_1, p_2, p_3 \ldots$ are proportions based on different numbers of observations $n_1, n_2, n_3 \ldots$ and if the n's differ widely, then the variances of the corresponding angular values may be so unequal, because of the variation of the n's, that the advantages of the transformation are lost almost entirely. Similarly the advantages of the inverse sine transformation may be lost to a large extent when the data to be analyzed is subject to other forms of variability in addition to sampling variation. In this case the variance of an observed proportion p is given by $\frac{P(1-P)}{n} + \sigma_e^2$, where the first term represents simple sampling variation and the second term the extraneous variation.

1.8.6 <u>A method of discovering Bartlett's adjustment</u>:

Bartlett (1936) gives his adjustment $\varphi(0) = \varphi(\frac{1}{4n})$, $\varphi(1) = \pi - \varphi(0)$. Without any justification Eisenhart (1947) therefore suggests the following:

Let the values of $\varphi(p) = 2 \arcsin\sqrt{p}$ where $p = \frac{x}{n}$ (x, 0, 1, 2 ... n), be plotted on the arithmetic scale of arithmetic-probability paper against the cumulative probability $\sum_{r=0}^{x} \frac{n!}{r!(n-r)!} P^r(1-P)^{n-r}$, plotted on the probability scale for a particular value of n and a selected value of P. If $n \geq 10$ and $.05 \leq P \leq .95$ it will be found that the plotted points for $2 \leq x \leq n-2$ are nearly linear, that the points for x = 1 and x = n-1 depart slightly from linearity, but that the points for x = 0 and x = n show considerable deviations. If the values $\varphi(0) = 0$ and $\varphi(1) = \pi$ are replaced by scores $\varphi(0) = a_n$ and $\varphi(1) = b_n$ such that the corresponding plotted points fall on a gently curving extrapolation of the curve through the plotted points for $2 \leqq x \leqq n-2$, it will be found that $a_n = 2 \arcsin \sqrt{\frac{1}{4n}}$ approximately.

$$b_n = \pi - a_n \quad \text{approximately.}$$

The foregoing suggests that a better adjustment than Bartlett's could be obtained by extending in both directions the line of best fit through the points for $2 \leqq x \leqq n-2$ and using this line to define the scores for $x = 0, 1, n-1$, and n. Eisenhart states that this was tried and some improvement was noted, but for $n \geqq 10$ and $.05 \leqq P \leqq .95$ the improvement did not appear to be worth the trouble.

1.9 <u>Transformation of data from entomological field experiments</u>: Beall (1942), in a study of experimental results from seven field experiments on the control of insects, shows that the variance as a function of the mean is not exactly $\sigma^2 = M$ as in the Poisson case but is better expressed as $\sigma^2 = M + kM^2$. The experiments are particularly designed so that the constant k can be estimated from the data. The transformation applied is $x' = k^{-\frac{1}{2}} \sinh^{-1}(kx)^{\frac{1}{2}}$, where k is a positive constant and x an observation. The data was put in a form for which the standard deviation approached a constant independent of the mean. Practically, the transformation gave good results so that the analysis of variance could be applied. From the analysis of the transformed data, the results were found to differ markedly from those which would have been obtained from the original data.

1.10 <u>Transformations related to the angular and the square root</u>: Freeman and Tukey (1950) report on an empirical study of a number of approximations to the normal deviate K exceeded with the same probability as the number of successes x from n in a Binomial distribution with expectation np, which is defined by: $\frac{1}{2\pi}\int_{-\infty}^{K} e^{-\frac{1}{2}t^2} dt = \text{prob.}\{x \leq k \mid \text{Binomial}, n, p\}$.

They also describe an angular and a square root transformation which yield better stabilizations of the variance.

The approximations to K in terms of n, p, k are given as:

$$N = 2(\sqrt{(k+1)q} - \sqrt{(n-k)p}) \quad .$$

$$N^{+} = N + \frac{N+2p-1}{12\sqrt{E}} \qquad E = \text{lesser of } np \text{ and } nq.$$

$$N^{*} = N + \frac{(N-2)(N+2)}{12}\left(\frac{1}{\sqrt{np+1}} - \frac{1}{\sqrt{nq+1}}\right) \quad .$$

$$N^{**} = N^{*} + \frac{N^{*}+2p-1}{12\sqrt{E}} \qquad (E \text{ as above}) \quad .$$

For variance stabilization, the averaged angular transformation, $\sin^{-1}\sqrt{\frac{x}{n+1}} + \sin^{-1}\sqrt{\frac{x+1}{n+1}}$ has a variance within ±6% of the quantities $\frac{1}{n+\frac{1}{2}}$ (angles in radian), $\frac{821}{n+\frac{1}{2}}$ (angles in degrees), for almost all cases where $np \geq 1$. In the Poisson case this amounts to using $\sqrt{x} + \sqrt{x+1}$ as having variance 1.

Fig. VI shows a comparison between the variances of the following five transformations:

$$\sqrt{x},\ \sqrt{x+1},\ \sqrt{x+\tfrac{1}{2}},\ \sqrt{x+\tfrac{3}{8}},\ \sqrt{x} + \sqrt{x+1}\ .$$

The figure indicates that for $m \lesseqgtr 3$, the various transformations may be arranged from the worst to the best as follows:

$$\sqrt{x},\ \sqrt{x+1},\ \sqrt{x+\tfrac{1}{2}},\ \sqrt{x+\tfrac{3}{8}},\ \sqrt{x} + \sqrt{x+1}\ .$$

2. Theoretical treatment of the stabilization of variance

2.1 The analysis of variance when experimental errors follow the Poisson or Binomial law:

2.1.1 Poisson case: The first step in an exact statistical analysis of the results of any field experiment is to specify in mathematical terms (1) how the expected values of the yield of the plots are obtained in terms of unknown parameters representing the treatment and block (or row and column) effects, and (2) how the observed values of the yields of

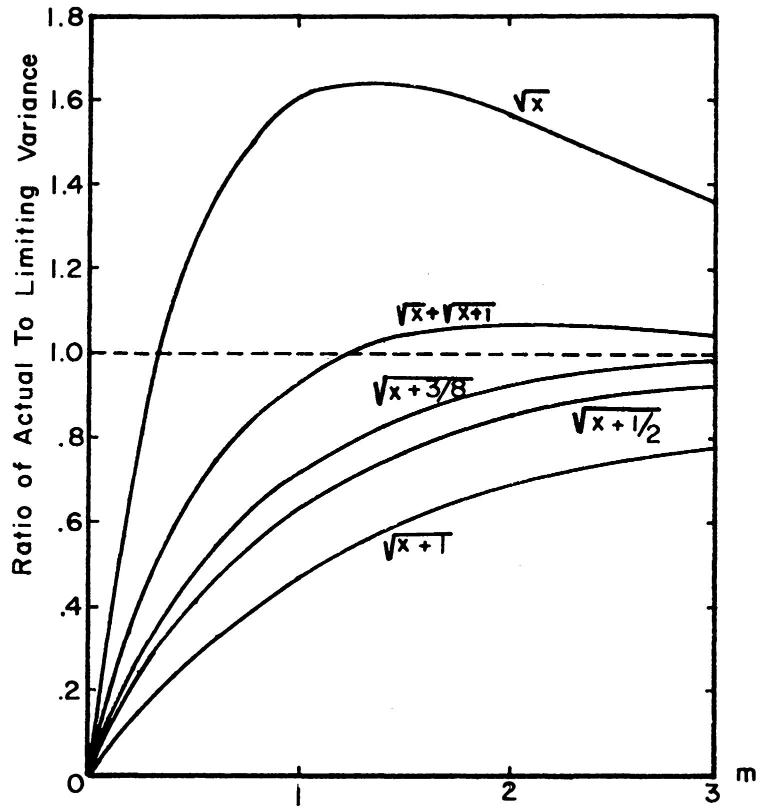

Fig. VI Stabilization of Poisson Variance the plots vary about the expected values. We assume here that variations follow the Poisson law.

Proceeding from the assumption that treatment and block (or row and column) effects are additive, we can specify the expected yield m_i of the ith plot, which receives the tth treatment and occurs in the rth row and the cth column as:

$$m_i = G + T_t + R_r + C_c \quad , \tag{49}$$

where G is a parameter representing the average level of the yield in the experiment and T_t, R_r, C_c represent the respective treatment, row, and column effects, to which the plot corresponds. We may put as usual,

$$\sum_t T_t = \sum_r R_r = \sum_c C_c = 0 \quad . \tag{50}$$

If the errors are normally and independently distributed with equal variances, this will lead to simple equations of estimation. In our case the probability of obtaining a given set of plot yields x_i with expectations m_i may be written as $\prod_i \frac{e^{-m_i} m_i^{x_i}}{x_i!}$. Thus L, the logarithm of the likelihood is given by:

$$L = \sum_i (x_i \log m_i - m_i) - \sum_i \log x_i! \quad , \tag{51}$$

Hence the maximum likelihood equation of estimation for any parameter θ assumes the form:

$$\frac{\Sigma(x_i - m_i)}{m_i} \cdot \frac{\partial m_i}{\partial \theta} = 0 \quad , \tag{52}$$

where the summation extends over all plots whose expectations involve θ. The function $\frac{\partial m_i}{\partial \theta}$ will usually involve a number of parameters. Since the specification of row, column and treatment effects in a 6 x 6 Latin square requires 16 independent parameters, the solution of these equations may be expected to be laborious. The problem of obtaining exact tests of significance is also difficult. The method of maximum likelihood provides estimates of the asymptotic variances and covariances of the treatment constants, which under certain conditions can be assumed to be asymptotically normally distributed. These asymptotic properties will hold if there is sufficient replication but will not apply in general for small samples. These remarks show that the exact solution is somewhat too complicated for frequent use. The difficulty arises principally because the typical equation of estimation consists of a weighted sum of the deviations of the observed from the expected values, the weights being $\frac{1}{m_i} \frac{\partial m_i}{\partial \theta}$. The factor $\frac{1}{m_i}$ would be removed from the weight function by using a square root transformation of the observations. For a Latin square, this prediction formula is written:

$$\sqrt{m_i} = \alpha_i = G + T_t + R_r + C_c \quad , \tag{53}$$

where

$$\sum_t T_t = \sum_r R_r = \sum_c C_c = 0 \quad . \tag{54}$$

To find the maximum value of (51) subject to the restrictions (54), we may use the method of undetermined multipliers, maximizing with respect to the parameters

$$L + \lambda(\sum_t T_t) + \mu(\sum_r R_r) + \nu(\sum_c C_c) \quad . \tag{55}$$

The equation of estimation for a typical treatment constant T_t becomes,

$$\sum \frac{(x_i - m_i)}{m_i} \frac{dm_i}{d\alpha_i} \frac{\partial \alpha_i}{\partial T_t} + \lambda = 0 \quad . \tag{56}$$

Since $\frac{dm_i}{d\alpha_i} = 2\sqrt{m_i}$, and $\frac{\partial \alpha_i}{\partial T_t} = 1$, (56) becomes:

$$\sum \frac{2(x_i - m_i)}{\sqrt{m_i}} + \lambda = 0 \quad . \tag{57}$$

The summation in (57) is extended over all plots receiving the treatment. If $a_i = \sqrt{x_i}$, then we might write:

$$x_i - m_i = a_i^2 - \alpha_i^2 = (a_i - \alpha_i)\frac{dm_i}{d\alpha_i} + \frac{1}{2!}(a_i - \alpha_i)^2 \frac{d^2 m_i}{d\alpha_i^2} \quad . \tag{58}$$

If m_i is reasonably large, only the first term on the right-hand side need be retained. When m_i is small, we may use, instead of the exact square root, a quantity a_i' defined so that

$$\begin{aligned} x_i - m_i &= (a_i' - \alpha_i)\frac{dm_i}{d\alpha_i} \\ &= 2\sqrt{m_i}\,(a_i' - \alpha_i) \quad . \end{aligned} \tag{59}$$

Thus if the analysis is performed on the quantities a_i' instead of on the original data, equation (57) becomes:

$$\sum_{T_t} 4(a_i' - \alpha_i) + \lambda = 0 \quad . \tag{60}$$

On substituting the expectations for α_i from (53) and using (54), we obtain

$$\sum_{T_t} 4(a_i' - G - T_t) + \lambda = 0 \tag{61}$$

because $\sum_{T_t} R_r = \sum_r R_r = 0$, $\sum_{T_t} C_c = \sum_c C_c = 0$. (The treatment T_t occurs once in each row and once in each column.) The corresponding equation for G is

$$\sum_i 4(a_i' - G) = 0 \quad , \tag{62}$$

so that G is the general mean of the quantities a_i'. By summing equation (61) over all treatments and comparing the total with (62) we find $\lambda = 0$. Hence T_t is the difference between the mean yield of a' over all plots receiving T_t and the general mean of a'. With this scale the simplicity of the normal theory equations has apparently been recovered. Actually the quantities a' are not known exactly since we have by (59)

$$a' = \alpha + \frac{x-m}{2\sqrt{m}} = \frac{1}{2}(\alpha + \frac{x}{\alpha}) \quad , \tag{63}$$

where α is the expected value of $\sqrt{x}$. However, this process provides a means of successively approximating the maximum likelihood solution by starting with approximations to the quantities α, then constructing the a''s and solving for the unknown constants and hence obtaining second approximations to the expected values. The close relation of a' to $\sqrt{x}$ is seen by remembering one of the common rules for finding square roots. This consists in guessing an approximate root α, dividing x by the approximate root and taking the mean of the approximate root α and the resulting quotient $\frac{x}{\alpha}$. The suitability of the linear prediction formula in square roots must be considered in any example in which the above analysis is being employed.

In this connection it should be noted that an approximate "goodness of fit" test of the validity of the implied assumptions may be obtained. Since the quantities a_i' enter into the equations of estimation with weight 4, the quantity $4\sum_i(a_i'-\alpha_i)^2$ is distributed approximately as χ^2 with the number of degrees of freedom in the error term of the analysis of variance. Some idea of the closeness of the approximation may be gathered by considering the simplest case in which only the mean yield is being estimated. In this case the observed value x are assumed to be drawn from the same Poisson distribution, and the sufficient statistic for the mean G is known to be $\sum \frac{x_i}{n}$. Since, however, the prediction formula is invariant under the transformation, and since the maximum likelihood solution is invariant to changes of scale, the mean value α of a' must be exactly $\sqrt{\frac{\sum x_i}{n}}$. Thus $a'-\alpha = \frac{1}{2}(\frac{x}{\alpha} - \alpha)$

and

$$\sum 4(a'-\alpha)^2 = \sum\frac{(x-\alpha^2)^2}{\alpha^2} = \sum\frac{(x-\bar{x})^2}{\bar{x}} \quad . \tag{64}$$

The usual χ^2 test for examining whether a set of values of x come from the same Poisson distribution may be assumed.

A high value of χ^2 means either that the prediction formula is not satisfactory or that the experimental errors are higher than the Poisson distribution indicates, or that both causes are operating.

2.1.2 <u>Binomial case</u>: In this case the yields are obtained by examining a constant number n per plot and noting those which possess a certain attribute (e.g., plants which are diseased). Experimental variation of the observed fraction p possessing the attribute about the expected fraction P. The proportion P is specified in terms of unknown parameters representing the treatment and soil effects.

If r_i is the number possessing the attribute on a typical plot, so that $p_i = \frac{r_i}{n}$, the likelihood function takes the form,

$$\prod_i \frac{n!}{r!(n-r)!} P_i^{r_i} Q_i^{n-r_i} \quad . \tag{65}$$

Hence

$$L = \sum_i [r_i \log P_i + (n-r_i) \log Q_i] \quad . \tag{66}$$

The equation of estimation for a typical parameter θ is

$$\Sigma \frac{n}{P_i Q_i} (p_i - P_i) \frac{\partial P_i}{\partial \theta} = 0 \quad , \tag{67}$$

where the summation is over all plots whose expectations involve θ.

As in the Poisson case, an exact solution is laborious because of the weights $\frac{n}{P_i Q_i} \frac{\partial P_i}{\partial \theta}$. The unequal weighting may be removed by transforming to the variate $\alpha_i = \sin^{-1}\sqrt{P_i}$ and assuming that the prediction formula is linear in the transformed scale. For a Latin square the prediction formula is assumed to be:

$$\alpha_i = G + T_t + R_r + C_c \quad , \tag{68}$$

where the ith plot receives treatment t and lies in the rth row and cth column. Further

$$\sum_t T_t = \sum_r R_r = \sum_c C_c = 0 \quad . \tag{69}$$

Since $\frac{dP_i}{d\alpha_i} = 2\sqrt{P_i Q_i}$, we introduce a set of variates a_i' so that on each plot,

$$\begin{aligned} p_i - P_i &= (a_i' - \alpha_i)\frac{dP_i}{d\alpha_i}\frac{\partial \alpha_i}{\partial \theta} \\ &= 2\sqrt{P_i Q_i}\,(a_i' - \alpha_i) \quad . \end{aligned} \tag{70}$$

With these substitutions the equation of estimation for T_t, for instance, becomes

$$\sum_{T_t} 4n(a_i' - \alpha_i) + \lambda = 0 \quad , \tag{71}$$

where as before, λ is an undetermined multiplier. The remainder of the solution proceeds exactly as in the Poisson case, T_t being found to be the difference between the mean value of a_i' over all plots receiving this treatment and the general mean of a_i'. A χ^2 test may be made with $\Sigma\, 4n(a_i'-\alpha_i)^2$. From (70) we have:

$$a_i' = \alpha_i + \frac{1}{2\sqrt{P_iQ_i}}\,(p_i-P_i)$$

$$= \alpha_i + \frac{1}{2\sqrt{P_iQ_i}}\,(Q_i-q_i)$$

$$= \alpha_i + \frac{1}{2}\cot\alpha_i - q_i\,\text{cosec}(2\alpha_i) \quad , \qquad (72)$$

where q_i is the observed fraction which does not possess the attribute.

The calculation of approximations to a_i' thus involves finding a predicted value α_i from the treatment and block (or row and column) means, and using equation (72). Tables [Fisher and Yates (1953)] of the values of $\sin^{-1}\sqrt{P_i}$, $\alpha_i + \frac{1}{2}\cot\alpha_i$, and $\text{cosec}(2\alpha_i)$ have been prepared to facilitate the computations. It should be noted that these tables are in degrees, whereas the above equations assume that α_i is measured in radians. In degrees, equation (71) above becomes

$$\sum_{T_t} \frac{\pi^2 n}{8100}\,(a_i'-\alpha_i)^2 = 0 \quad , \qquad (73)$$

while

$$a_i' = \alpha_i + \frac{180}{\pi}\,\{\frac{1}{2}\cot\alpha_i - q_i\,\text{cosec}(2\alpha_i)\} \quad . \qquad (74)$$

It must be emphasized that the solutions for the Poisson and Binomial situations given above apply to the case where the whole of the experimental error variation is of the Poisson or Binomial type. The methods are therefore likely to be useful in practice only where the experimental conditions have been carefully controlled, or where the data are derived from such small numbers that the Poisson or Binomial variation is much larger than any extraneous variation. The χ^2 test is helpful in deciding whether this assumption is justified.

2.2 Asymptotic theory and applications:

2.2.1 Introduction: Curtiss (1943) provided a general theoretical approach to the problem of transformations. In the framework of this theory he discussed in particular the square root and inverse sine transformations and also several logarithmic transformations.

In general we can state the problem of stabilizing a variance functionally related to a mean as follows:

Suppose x is a variate whose mean $\mu = E(x)$ is a real variable with a range s of possible values and whose standard deviation $\sigma = \sigma_x = \sigma(\mu)$ is a function of μ not identically constant. We desire to find a function $T = f(X)$ such that both $f(X)$ and σ_T^2 are functionally independent of μ for μ on s.

Curtiss first of all criticizes any transformation which purports to produce an absolutely constant variance. Taking for instance a Poisson variate X and assuming the identity:

$$E[\{f(x) - E[f(x)]^2\}]^2 \underset{\mu}{\equiv} c \quad ,$$

or equivalently,

$$E\{[f(x)]^2\} \underset{\mu}{\equiv} c + \{E[f(x)]\}^2 \quad ,$$

we have:

$$\sum_{k=0}^{\infty}[f(k)]^2 \frac{e^{-\mu}\mu^k}{k!} \underset{\mu}{\equiv} c + [\sum_{k=0}^{\infty}[f(k)] \frac{e^{-\mu}\mu^k}{k!}]^2 \quad , \quad \mu > 0 \quad .$$

We need only equate the coefficients of the zero-th power of μ on each side to find that $[f(0)]^2 = c + [f)0)]^2$ which implies that $c = 0$ and hence that $f(0) = f(1) = f(2) \ldots$, so that the solution is trivial, namely $T \equiv$ Const. As to the problem of choosing $T = f(X)$ so that its distribution is exactly normal, we can observe immediately that a single valued function $f(X)$ will never transform a variate X with a discrete distribution into a variate with a continuous one. On the other hand, any variate X with a continuous distribution function $F(x)$ can be transformed into a normally distributed variate T by the transformation $T = f(X)$ defined by the equation $F(x) = \int_{-\infty}^{T} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt$. However, aside from the practical difficulty of solving this equation for T, the resulting function $T = f(X)$ will not generally be functionally independent of the mean of X.

This consideration leads Curtiss to seek asymptotic solutions to the problem of normalization and stabilization.

In the following asymptotic theorems we suppose that the distribution of X depends on a parameter n which is to tend somehow to infinity. The mean $\mu = \mu_n$ and the range s_n of X will in general depend on n (We include the case μ_n = Const. for all n.) and perhaps will depend also on some further independent parameters which we shall denote collectively by θ, with range Σ. We shall seek a transofrmation $T = f(X)$ in which $f(X)$ is functionally independent of μ and of θ for μ on s_n and θ on Σ, and such that the distribution of $f(X) - f(\mu_n)$ tends to normality as $n \to \infty$, while $\lim_{n\to\infty} \sigma_T^2 = c^2$, where c^2 is an absolute constant.

2.2.2 <u>Theorem</u>: Let $\Psi_n(X)$ be a non-negative function of x and n, defined almost everywhere and integrable with respect to x over any finite interval of the x-axis for each $n > 0$.

Let

$$T = f(X) = \int_a^X \Psi_n(x)\,dx \quad , \tag{75}$$

where a is an arbitrary constant. Let $F_n(y)$ be the distribution function (d.f.) of the variate $Y = (X - \mu_n)\Psi_n(\mu_n)$. Suppose further that a continuous d.f. exists such that $\lim_{n\to\infty} F_n(y) = F(y)$ for all values of y. Then either one of the following two conditions is a sufficient condition for the distribution function $H_n(w)$ of the variate $W = f(X) - f(\mu_n)$ to tend uniformly to $F(w)$, $-\infty < w < +\infty$.

a) To each w for which $0 < F(w) < 1$, there corresponds for all n sufficiently large at least one root $x = x_n$ of the equation

$$\int_{\mu_n}^{x} \Psi_n(u)\,du = w \quad , \tag{76}$$

and this root x_n has the property that

$$\lim_{n\to\infty} (x_n - \mu_n)\Psi_n(\mu_n) = w \quad . \tag{77}$$

b) For all n sufficiently large, $\Psi_n(\mu_n) > 0$, and $\lim_{n\to\infty} q_n(w) = 1$ uniformly in any closed finite subinterval of the open interval defined by $0 < F(w) < 1$, where

$$q\ (w) = \frac{\Psi_n(w[\Psi_n(\mu_n)]^{-1} + \mu_n)}{\Psi_n(\mu_n)} \quad . \qquad (78$$

To prove this theorem we first suppose that condition (a) is satisfied. Let w_1 and w_2 be the end points of the open interval (possibly infinite) defined by $0 < F(w) < 1$. If w lies in this interval, and if n is large enough for the root x_n in (76) to exist, then from the monotonic character of $\int_{\mu_n}^{X} \Psi_n(x)\,dx$ we can infer that

$$H_n(w) = P[f(X)-f(\mu_n) \leqq w] = P[\int_{\mu_n}^{X} \Psi_n(x)\,dx \leqq w]$$

$$= P[\int_{\mu_n}^{X} \Psi_n(x)\,dx \leqq \int_{\mu_n}^{x_n} \Psi_n(x)\,dx]$$

$$= P[X \leqq x_n] = P[Y \leqq (x_n-\mu_n)\Psi_n(\mu_n)]$$

$$= F_n[(x_n-\mu_n)\Psi_n(\mu_n)] \quad . \tag{79}$$

Since F(w) is continuous, $\lim_{n\to\infty} F_n(w) = F(w)$ uniformly on any finite or infinite interval of values of w. Therefore $\lim_{n\to\infty} F_n(w_n) = F(w)$ if $\lim_{n\to\infty} w_n = w$. Thus from (77) and (79), we find that

$$\lim_{n\to\infty} H_n(w) = F(w) \qquad \text{for } w_1 < w < w_2 \quad . \tag{80}$$

If $w' \leqq w_1$, and $w_1 < w'' < w_2$, then $0 \leqq H_n(w') \leqq H_n(w'') = F(w'') + [H_n(w'')-F(w'')]$. We can make the right-hand member of this relation less than any given positive number ε by first choosing w" so that $f(w'') < \frac{1}{2}\varepsilon$ (it will be remembered that $F(w)$ is a continuous d.f., and $F(w_1) = 0$) and then choosing n so large that the quantity in square brackets is also less than $\frac{1}{2}\varepsilon$ in absolute value. Thus $\lim_{n\to\infty} H_n(w') = 0$. Similarly if $w' \geqq w_2$ we can show that $\lim_{n\to\infty} H_n(w') = 1$. Hence $\lim_{n\to\infty} H_n(w) = F(x)$ for all w, and it follows that the limit holds uniformly on any finite or infinite interval of values of w.

We show now that condition (a) in the theorem is a consequence of condition (b). The result follows at once from the following lemma.

<u>Lemma</u>: If $\gamma_n(w)$ is a non-negative function, integrable over any finite interval of w, and if $\lim_{n\to\infty} \gamma_n(w) = 1$ uniformly in any finite closed subinterval of an interval $w_1 < w < w_2$, then for every value of w in this interval there exists for all n sufficiently large a solution $y = y_n$ of the equation:

$$\int_0^y \gamma_n(z)\,dz = w \quad , \tag{81}$$

and the solution y_n has the property that $\lim_{n\to\infty} y_n = w$.

It is clear that if w satisfies the inequality $w_1 < w < w_2$, and if $\eta > 0$ be chosen so that $w_1 < w-\eta < w+\eta < w_2$, then for all n sufficiently large,

$$\int_0^{w-\eta} \gamma_n(z)\,dz \leq w \leq \int_0^{w+\eta} \gamma_n(z)\,dz \quad ,$$

(because we know that $\gamma_n(z)$ is a non-negative and integrable and $\lim_{n\to\infty} \gamma_n(w) = 1$ in any finite closed subinterval of w). Thus for each n sufficiently large, there exist a root y_n of the equation (81), and furthermore, this root satisfies the inequality $w-\eta \leq y_n \leq w+\eta$. Since η is arbitrarily small, $\lim_{n\to\infty} y_n = w$ and the proof of the lemma is complete.

To apply the lemma we make the change of variables,

$$z = (u-\mu_n)\Psi_n(\mu_n) \quad \text{or} \quad u = [z/\Psi_n(\mu_n)] + \mu_n \quad ,$$

$$\Psi_n(u) = \Psi_n[z[\Psi_n(\mu_n)]^{-1} + \mu_n] \quad .$$

Substituting in the integral (76), it becomes

$$\int_0^{y} q_n(z)\,dz \quad , \tag{82}$$

where $y = (x-\mu_n)\Psi_n(\mu_n)$, and the conclusion that (a) is implied by (b) now follows at once.

2.2.3 <u>Theorem</u>: Let T (or f(X), Y, $F_n(y)$, and F(y) be defined as in theorem (2.2.2). Let the mean and variance of the distribution defined by F(y) exist and have respective values zero and c^2. Then the following three conditions, taken together:

(i) $E(Y^2)$ exists for $n > 0$, and $\lim_{n\to\infty} E(Y^2) = c^2$,

(ii) Condition (b) of theorem (2.2.2) holds ,

(iii) $f(Y[\Psi_n(\mu_n)]^{-1}+\mu_n) - f(\mu_n) = O|Y|$ uniformly in n as $|Y| \to \infty$, are sufficient that,

$$\lim_{n\to\infty} [E(T) - f(\mu_n)] = 0 \quad , \tag{83}$$

$$\lim_{n\to\infty} \sigma_T^2 = c^2 \quad . \tag{84}$$

The proof is divided into four parts:

a) As a preliminary step in the proof, we observe that condition (i) and the relations, $\lim_{n\to\infty} F_n(y)=F(y)$, $c^2=\int_{-\infty}^{+\infty} y^2 dF(y)$, imply that the improper integral $\int_{-\infty}^{+\infty} y^2 dF_n(y)$ converges uniformly in n for $n > 0$. As the integrand is positive, the following result is equivalent to the uniform convergence of the integral; for every $\varepsilon > 0$, there exist numbers A_1 and A_2, $A_1 < A_2$, such that for all n sufficiently large

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2 d\, F_n(y) < \varepsilon \quad . \tag{85}$$

To prove (85) we write

$$\left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right) y^2 d\, F_n(y) = [E(Y^2)-c^2]+\left[\int_{A_1}^{A_2} y^2 dF(y) - \int_{A_1}^{A_2} y^2 dF_n(y)\right]$$
$$+ \left[c^2 - \int_{A_1}^{A_2} y^2 dF(y)\right] \quad .$$

We first choose A_1, A_2 so that the last bracket approaches zero as $n \to \infty$, and then Helly-Bray theorem (G. C. Evans, 1927) states that the second bracket also approaches zero as $n \to \infty$. So for all n sufficiently large, the sum of the first two brackets is in absolute value $< \frac{1}{2}\varepsilon$.

It is important to notice that we can choose A_1 and A_2 in the above demonstration so that $A_1 > w_1$, $A_2 < w_2$, where w_1 and w_2 are as usual the endpoints of the interval defined by $0 < f(w) < 1$.

(b) We try now to express (83) and (84) in terms of the variate $W = f(X) - f(\mu_n)$ and then to establish a relation between (W) and $Y = (X-\mu_n)\Psi_n(\mu_n)$. For we need merely to notice that

$$W = f(X)-f(\mu_n)=\int_a^X \Psi_n(x)\,dx - \int_a^{\mu_n} \Psi_n(x)\,dx = \int_{\mu_n}^X \Psi_n(x)\,dx$$

and then to make a change of variable similar to the one used to derive (82). Now we can write

$$W = \int_{\mu_n}^X \Psi_n(x)\,dx = \int_0^Y q_n(w)\,dw = Q_n(Y) \quad , \tag{86}$$

where $q_n(w)$ is given by (78).

In terms of W, (83) and (84) become respectively:

$$\lim_{n\to\infty} E[f(X)-f(\mu_n)] = \lim E(W) = 0 \quad (87)$$

$$\lim_{n\to\infty} \{E(W^2)-[E(W)]^2\} = \sigma_W^2 = \sigma_T^2 = c^2 \quad . \quad (88)$$

We need to establish now (87) and (88) instead of (83) and (84).

(c) For (87), condition (ii) states that for all n sufficiently large $\Psi_n(\mu_n) > 0$ and $\lim_{n\to\infty} q_n(w) = 1$ uniformly in any closed finite subinterval of the open interval defined by $0 < F(w) < 1$. This implies that

$$\lim_{n\to\infty} Q_n(y) = \lim_{n\to\infty}\int_0^y q_n(w)\,dw = \int_0^y \lim_{n\to\infty} q_n(w)\,dw = \int_0^y dw = y$$

uniformly in any finite closed subinterval of the interval $w_1 < y < w_2$; condition (iii) states that $f(Y[\Psi_n(\mu_n)]^{-1}+\mu_n) - f(\mu_n) = O|Y|$ uniformly in n as $|Y| \to \infty$. Since $Y = (X-\mu_n)\Psi_n(\mu_n)$ this relation may be written as $f(X) - f(\mu_n) = O|Y|$ as $|Y| \to \infty$; this implies that there exists a constant M such that $|f(x)-f(\mu_n)|=|W|=|Q_n(y)|\leq M|y|$ for all n. Moreover if $E(Y^2)$ exists, so will $E(Y)$.

Now $$E(W)=\int_{-\infty}^{+\infty} Q_n(y)\,dF_n(y)=\int_{-\infty}^{\infty} Q_n(y)\,dF_n(y)-\int_{-\infty}^{+\infty} y\,dF_n(y) \qquad (E(y)=0)$$

$$= \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)[Q_n(y)-y]\,dF_n(y)+\int_{A_1}^{A_2}[Q_n(y)-y]\,dF_n(y) \quad ,$$

where $w_1 < A_1 < A_2 < w_2$. Therefore

$$|E(W)| \leq \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)(M+1)|y|\,d\,F_n(y) + \int_{A_1}^{A_2} |Q_n(y)-y|\,d\,F_n(y) .$$

From the uniform convergence of $\int_{-\infty}^{+\infty} y^2\,d\,F_n(y)$, proved in part (a) we can conclude that the first pair of integrals can be made less than an arbitrary $\frac{1}{2}\varepsilon > 0$, by proper choice of A_1, A_2. The third integral approaches zero by the general Helly-Bray theorem (We have shown that $\lim_{n\to\infty} Q_n(y) = y$.) and so becomes less than $\frac{1}{2}\varepsilon$ for all n sufficiently large. Thus $|E(W)| \leq \varepsilon$ for all n sufficiently large and we have established (87).

(d) To show that (88) is true, we have merely to prove that $\lim_{n\to\infty} E(W^2) = c^2$. Since $E(Y^2) = \int_{-\infty}^{+\infty} y^2\,d\,F_n(y)$, we may write:

$$E(W^2)-c^2 = \int_{-\infty}^{+\infty} \{[Q_n(y)]^2-y^2\}dF_n(y)+[E(Y^2)-c^2] .$$

But

$$\int_{-\infty}^{+\infty} \{[Q_n(y)]^2-y^2\}dF_n(y) = \left(\int_{-\infty}^{+A_1} + \int_{A_2}^{\infty}\right)\{[Q_n(y)]^2-y^2\}dF_n(y)$$

$$+ \int_{A_1}^{A_2} \{[Q_n(y)]^2-y^2\}dF_n(y)$$

$$\leq \left(\int_{-\infty}^{A_1} + \int_{A_2}^{\infty}\right)(M+1)y^2dF_n(y) + \int_{A_1}^{A_1} \{[Q_n(y)]^2-y^2\}dF_n(y) .$$

Since $\int_{-\infty}^{+\infty} y^2 d\, F_n(y)$ is uniformly convergent we can choose A_1 and A_2 such that the first part $< \frac{1}{2}\varepsilon$ $(\varepsilon > 0)$. Again by the general Helly-Bray theorem, the second part approaches zero, and so for n sufficiently large, it can be made $< \frac{1}{2}\varepsilon$. Hence

$$\int_{-\infty}^{+\infty} \{[Q_n(y)]^2 - y^2\} dF_n(y)$$

approaches zero, as also does $[E(Y^2) - c^2]$ by condition (i) so that $E(W^2) - c^2 \to 0$, or

$$E(W^2) \to c^2 \quad \text{as } n \to \infty \; .$$

The proof of the theorem is now established.

2.2.4 <u>Theorem</u>: Let the distribution of a variate Y depend upon a parameter n. Let $F_n(y)$ be the d.f. of Y, and let $F(y)$ be a continuous d.f. with the property that $\lim_{n\to\infty} F_n(y) = F(y)$. Let a_n be a function of n such that $\lim_{n\to\infty} a_n = a \neq 0$. Then the distribution function of the variate $z = a_n Y$ tends as $n \to \infty$ to the distribution function $F(\frac{z}{a})$ if $a > 0$, and to the d.f. $1 - F(\frac{z}{a})$ if $a < 0$. If the variance of Y exists and tends to c^2 as $n \to \infty$, then the variance of $a_n Y$ tends to $a^2 c^2$ as $n \to \infty$.

No proof is provided for this theorem, but we know in a special case that if $F(y)$ is the d.f. of a reduced normal distribution, i.e., $F(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{1}{2}t^2} dt$, then $F(\frac{z}{a})$ is also the d.f. of a normal distribution with mean zero and variance a^2. More generally, any affine transformation of a normal variate yields another normal variate.

2.3 Applications

2.3.1 Introduction: The theorems of the preceding section have the effect of referring the properties of the distribution of the transformation $T = f(X)$ of theorem I, back to those of the distribution of a related variate Y. In the applications given in the present section, we shall let $\Psi_n(\mu_n)$ be proportional to the reciprocal of the standard deviation of X. The preceding theorems state in this case that if the reduced or standardized distribution of X approaches a limiting form, then under certain circumstances the distribution of $f(X) - f(\mu_n)$ will approach a similar limiting form and σ_T^2 will approach a quantity independent at least of the parameter n. In the applications considered here, the reduced distribution of X will always approach the reduced normal distribution.

2.3.2 The square root transformation for a variate with a Poisson distribution: Let X have a Poisson distribution with parameter n. If α is an arbitrary constant, and if

$$T = f(X) = \begin{cases} \sqrt{x+\alpha} & x \geq -\alpha \\ 0 & x < -\alpha \end{cases} . \tag{89}$$

Then the distribution of $T - \sqrt{n+\alpha}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $\frac{1}{4}$, and $\lim_{n\to\infty} \sigma_T^2 = \frac{1}{4}$.

<u>Proof</u>: Here $\mu_n = n$, $\sigma_X = \sqrt{n}$, and it is well known that the distribution of the reduced variate $\frac{X-\mu_n}{\sqrt{n}}$ tends to the reduced normal distribution as $n \to \infty$. By theorem (2.2.3) the distribution of $Y = \frac{X-n}{2\sqrt{n+\alpha}} = \frac{1}{2}\sqrt{\frac{n}{n+\alpha}} \cdot \frac{X-n}{\sqrt{n}}$, will tend to normality as $n \to \infty$ and the variance of $Y = \frac{1}{4}\frac{n}{n+\alpha}\text{Var}(\frac{X-n}{\sqrt{n}}) \to \frac{1}{4}$ as $n \to \infty$. The mean of $Y = 0$.

Setting:

$$\Psi_n(X) = \begin{cases} \frac{1}{2\sqrt{X+\alpha}} & x > -\alpha \\ 0 & x \leq -\alpha \end{cases} . \tag{90}$$

We obtain from $T = f(X) = \int_{-\alpha}^{X} \Psi_n(x)\,dx = \int_{-\alpha}^{X} \frac{dx}{2\sqrt{x+\alpha}} = \sqrt{x+\alpha}$, the formula given in (89). Here $f(x) - f(\mu_n) = W = T - \sqrt{n+\alpha}$. To complete the proof of the above statement we must show that conditions (ii) and (iii) of the theorem (2.2.3) are satisfied. Assuming $n > -\alpha$, we have

$$q_n(w) = \frac{\Psi_n(w[\Psi_n(\mu_n)]^{-1}+\mu_n)}{\Psi_n(\mu_n)} = \Psi_n(2w\sqrt{n+\alpha}+n) \cdot 2\sqrt{n+\alpha}$$

$$= \frac{2\sqrt{n+\alpha}}{2\sqrt{(2w\sqrt{n+\alpha}+n)+\alpha}} = (1 + \frac{2w}{\sqrt{n+\alpha}})^{-\frac{1}{2}}$$

provided that $w > -\frac{1}{2}\sqrt{n+\alpha}$; otherwise $q_n(w) = 0$. Hence clearly (ii) is satisfied simply because $\Psi_n(\mu_n) > 0$ and $\lim_{n\to\infty} q_n(w) = 1$. Also

$$W = f(Y[\Psi_n(\mu_n)]^{-1}+\mu_n) - f(\mu_n) = f(2Y\sqrt{n+\alpha}+n) - \sqrt{n+\alpha}$$

$$= \begin{cases} \sqrt{2Y\sqrt{n+\alpha}+n+\alpha} - \sqrt{n+\alpha} & \text{when } Y > -\frac{1}{2}\sqrt{n+\alpha}, \\ -\sqrt{n+\alpha} & \text{when } Y \leqq -\frac{1}{2}\sqrt{n+\alpha}, \end{cases}$$

from which it follows at once that $|W| < 2|Y|$ when $Y < -\frac{1}{2}\sqrt{n+\alpha}$. When $Y > -\frac{1}{2}\sqrt{n+\alpha}$ we have $W > 0$, and we can write:

$$W = \sqrt{2Y\sqrt{n+\alpha}+n+\alpha} - \sqrt{n+\alpha} < 2|Y| \quad \text{or equivalently:}$$

$$\sqrt{2Y\sqrt{n+\alpha}+n+\alpha} < 2|Y| + \sqrt{n+\alpha} \quad \text{or} \quad 2Y\sqrt{n+\alpha}+(n+\alpha) < (2|Y|+\sqrt{n+\alpha})^2$$

$$\text{or} \quad \sqrt{n+\alpha}\,[2Y+\sqrt{n+\alpha}] < (2|Y|+\sqrt{n+\alpha})^2,$$

and this is intuitive since both $\sqrt{n+\alpha}$ and $2Y+\sqrt{n+\alpha}$ are less or equal to $2|Y|+\sqrt{n+\alpha}$. Hence condition (iii) is satisfied for all Y.

2.3.3 <u>The square root transformation for a variate with a Γ distribution</u>: Let X have a distribution whose density function is of the following type:

$$\varphi(x) = \begin{cases} 0 & x \leqq 0 , \\ K\, x^{\frac{1}{2}n-1}\, e^{-hx} & x \geqq 0 ,\quad h > 0 \end{cases} . \qquad (91)$$

If α is an arbitrary constant, and if

$$T = f(X) = \begin{cases} \sqrt{X+\alpha} & X \geqq -\alpha \\ 0 & X < -\alpha \end{cases} , \qquad (92)$$

then the distribution of $T - \sqrt{\frac{n}{2h} + \alpha}$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $\frac{1}{4h}$, and $\lim_{n\to\infty} \sigma_T^2 = \frac{1}{4h}$.

We know that the mean of X is $\mu_n = \frac{n}{2h}$ and that the standard deviation $\sigma_X = \frac{\sqrt{n}}{h\sqrt{2}}$. The distribution of the standardized variate tends to normality as $n \to \infty$, so that of the variate $Y = \frac{x-\mu_n}{2\sqrt{\mu_n+\alpha}} = \frac{1}{2}\sqrt{\frac{n}{nh+2h^2\alpha}} \cdot \frac{x-\mu_n}{\sqrt{\frac{\mu_n}{h}}}$ tends to normality and $\mathrm{Var}(Y) = \frac{1}{4}\,\frac{n}{nh+2h^2\alpha} \cdot \mathrm{Var}\left(\frac{x-\mu_n}{\sqrt{\mu_n/h}}\right)$ which tends to $\frac{1}{4h}$ as $n \to \infty$.

Setting $\Psi_n(x) = \begin{cases} \frac{1}{2\sqrt{x+\alpha}} & x > -\alpha \\ 0 & x \leqq -\alpha \end{cases}$,

we obtain

$$T = \int_{-\alpha}^{X} \Psi_n(X)\,dx = \int_{-\alpha}^{X} \frac{dx}{2\sqrt{x+\alpha}} = \sqrt{X+\alpha} \quad \text{as in (92).}$$

The work of verifying that the conditions of theorem II are satisfied is the same as in the case of the Poisson distribution treated above.

For example, if s^2 denotes the variance of a random sample of n+1 observations drawn from a normal parent distribution with variance σ^2, then it is well known that $(n+1)s^2$ is distributed according to (91) with $h = \frac{1}{2\sigma^2}$. We can thus deduce the further facts, also well known, that the distribution of $\sqrt{n+1}\, s - \sigma\sqrt{n}$ tends to normality, and that the variance of $s\sqrt{n+1}$ approaches the limiting value $\frac{1}{2}\sigma^2$. If n is an integer and $h = \frac{1}{2}$, the distribution defined by (92) is called a χ^2 distribution with n degrees of freedom and the variate is denoted by χ^2. Our conclusion in this case is that $\sqrt{2\chi^2} - \sqrt{2n}$ tends to be normally distributed with zero mean and unit variance. From this result and the fact that $\sqrt{2n-1} - \sqrt{2n} = O(n^{-\frac{1}{2}})$, it follows immediately that $\sqrt{2\chi^2} - \sqrt{2n-1}$ has the same limiting distribution as $\sqrt{2\chi^2} - \sqrt{2n}$. Hence we get the Fisher χ^2-transformation which we will discuss in a later section.

2.3.4 <u>The inverse sine transformation for a Binomial variate:</u> Let X be a Binomial variate with proportion P $(P = 0, \frac{1}{n}, \frac{2}{n}, \ldots \frac{n}{n})$. If α is an arbitrary constant, and if

$$T = f(X) = \begin{cases} \sqrt{n}\,\sin^{-1}\sqrt{X+\frac{\alpha}{n}} & \frac{-\alpha}{n} \leq X \leq 1-\frac{\alpha}{n} \\ 0 & X < \frac{-\alpha}{n} \text{ or } X > 1-\frac{\alpha}{n} \end{cases}, \quad (93)$$

where T is measured in radians, then the distribution of $T - \sqrt{n}\,\sin^{-1}\sqrt{P+\frac{\alpha}{n}}$ is asymptotically normal with zero mean and variance $\frac{1}{4}$.

Here let us take $\mu_n = P$ and $\sigma_X^2 = \frac{PQ}{n}$ where $Q = 1-P$. We know that the distribution of the reduced variate $\sqrt{n}\,(X-P)/\sqrt{PQ}$ will be asymptotically normal. Hence by theorem (2.2.4) the distribution of

$$Y = \frac{\sqrt{n}\,(X-P)}{2\sqrt{(P+\frac{\alpha}{n})(Q-\frac{\alpha}{n})}} = \frac{\sqrt{PQ}}{2\sqrt{(P+\frac{\alpha}{n})(Q-\frac{\alpha}{n})}} \cdot \frac{\sqrt{n}\,(X-P)}{\sqrt{PQ}} \quad (94)$$

will tend to normality with variance$=\frac{PQ}{4(P+\frac{\alpha}{n})(Q-\frac{\alpha}{n})}\,\mathrm{Var}(\frac{\sqrt{n}(X-P)}{\sqrt{PQ}}) \sim \frac{1}{4}$.

The variance of the limiting distribution F(Y) is also $\frac{1}{4}$.

Let

$$\Psi_n(x) = \begin{cases} \frac{\sqrt{n}}{2\sqrt{(x+\frac{\alpha}{n})(1-x-\frac{\alpha}{n})}} & -\frac{\alpha}{n} < x \leq 1-\frac{\alpha}{n} \\ 0 & \text{otherwise} \end{cases}. \quad (95)$$

Then

$$T = \int_{-\frac{\alpha}{n}}^{X} \Psi_n(x)\,dx = \int_{-\frac{\alpha}{n}}^{X} \frac{\sqrt{n}\,dx}{2\sqrt{(x+\frac{\alpha}{n})(1-x-\frac{\alpha}{n})}} .$$

Putting $x+\frac{\alpha}{n} = \sin^2 t$, $1-x-\frac{\alpha}{n} = \cos^2 t$, $dx = 2\sin t\ \cos t\, dt$, we get

$$T = \int_0^{\sin^{-1}\sqrt{X+\frac{\alpha}{n}}} \frac{\sqrt{n}\ 2\ \text{sint cost dt}}{2\ \text{sint cost}} = \sqrt{n}\ \sin^{-1}\sqrt{X+\frac{\alpha}{n}} \quad ,$$

which is the same as we assumed in (93).

In showing that conditions (ii) and (iii) of theorem (2.2.3) are satisfied, we shall assume for simplicity that $\alpha = 0$. We find that:

$$q_n(w) = \frac{\Psi_n(w[\Psi_n(\mu_n)]^{-1}+\mu_n)}{\Psi_n(\mu_n)} = \frac{\Psi_n(\frac{2\sqrt{PQ}\ w}{\sqrt{n}} + P)\,2\sqrt{PQ}}{\sqrt{n}}$$

$$= \frac{2\sqrt{PQ}}{\sqrt{n}} \cdot \frac{\sqrt{n}}{2\sqrt{(\frac{2w}{\sqrt{n}}\ \sqrt{PQ} + P)(1-P-\frac{2w}{\sqrt{n}}\ \sqrt{PQ})}}$$

$$= \frac{\sqrt{PQ}}{\sqrt{[PQ - \frac{2w}{\sqrt{n}}\ P\sqrt{PQ} + \frac{2w}{\sqrt{n}}\ Q\sqrt{PQ} - \frac{4PQw^2}{n}]}}$$

$$= (1 + 2w\,\frac{Q-P}{\sqrt{nPQ}} - \frac{4w^2}{n})^{-\frac{1}{2}} \quad ,$$

provided $-\frac{1}{2}\sqrt{\frac{nP}{Q}} < w < \frac{1}{2}\sqrt{\frac{nP}{Q}}$, otherwise $q_n(w) = 0$. When $n \to \infty$, $q_n(w) \to 1$ so (ii) is satisfied.

From the law of the mean in the form due to Schlomilch we have

$$W = f(Y[\Psi_n(\mu_n)]^{-1}+\mu_n) - f(\mu_n)$$

$$= f(Y\ 2\sqrt{\frac{PQ}{n}} + P) - \sqrt{n}\ \sin^{-1}\sqrt{P}$$

$$= \sqrt{n}\ \sin^{-1}\sqrt{P+2\sqrt{\frac{PQ}{n}}\cdot Y} - \sqrt{n}\ \sin^{-1}\sqrt{P}$$

$$= 2Y\left[\frac{1-\theta}{(1+2\theta\sqrt{\frac{Q}{nP}}\ Y)(1-2\theta Y\sqrt{\frac{P}{nQ}})}\right]^{\frac{1}{2}} , \qquad (96)$$

provided $0 < \theta < 1$, $-\frac{1}{2}\sqrt{\frac{nP}{Q}} < Y < \frac{1}{2}\sqrt{\frac{nQ}{P}}$.

The square of the denominator in (96) is

$$\frac{-4\theta^2}{n^2}Y^2 + 2Y[\theta\sqrt{\frac{Q}{nP}} - \theta\sqrt{\frac{P}{nQ}}] + 1 .$$

We are interested in the positive part of this quadratic expression when $-\frac{1}{2}\sqrt{\frac{nP}{Q}} < Y < \frac{1}{2}\sqrt{\frac{nQ}{P}}$. Since the coefficients of Y^2 is negative, the expression is convex in this interval and so must assume its least value in that range at one end or the other. But when $Y = -\frac{1}{2}\sqrt{\frac{nP}{Q}}$, we get $W = 2Y[\frac{1}{1+\theta\frac{P}{\sqrt{Q}}}]^{\frac{1}{2}}$, and when $Y = \frac{1}{2}\sqrt{\frac{nQ}{P}}$, $W = 2Y[\frac{1}{1+\theta\frac{Q}{\sqrt{P}}}]^{\frac{1}{2}}$. Since these are the maximum possible coefficients of 2Y, and since both of them are less than one, we deduce that W is always less than (2Y) in the range indicated. For values of Y outside the range, the second member of the expression for W, namely $-\sqrt{n}\ \sin^{-1}\sqrt{P}$, indicates that $W = 0(\sqrt{n}) = 0(Y)$, because we know that Y is always of the order of $\sqrt{n}$, ($Y = \frac{\sqrt{n}\,(X-P)}{2\sqrt{PQ}}$, X, P, Q are less than one). Hence (iii) is satisfied and the proof of the statement above is complete for $\alpha = 0$.

2.3.5 <u>Other transformations of a Binomial variate</u>: Let X be a Binomial variate with parameter P ($P = 0, \frac{1}{n}, \frac{2}{n} \dots \frac{n}{n}$).

(a) If

$$T = f(X) = \begin{cases} \sqrt{n}\,\sinh^{-1}\sqrt{X} = \sqrt{n}\,\log(\sqrt{X} + \sqrt{1+X}\,) & X \geqq 0 \\ 0 & X < 0 \end{cases},$$

then the distribution of $T - \sqrt{n}\,\sinh^{-1}\sqrt{P}$ tends as $n \to \infty$ to a normal distribution which has zero mean and variance $\frac{Q}{4+4P}$.

Here we have $f'(x) = \frac{\sqrt{n}}{2\sqrt{x^2+x}}$, $x > 0$.

Setting:

$$\Psi_n(x) = \begin{cases} \sqrt{n}/2\sqrt{x^2+x} & x > 0 \\ 0 & x \leqq 0 \end{cases},$$

we obtain

$$Y = (X-P)\Psi_n(P) = \frac{\sqrt{n}(X-P)}{\sqrt{PQ}} \cdot \frac{\sqrt{PQ}}{2\sqrt{P}\,\sqrt{1+P}} = \frac{\sqrt{n}(X-P)}{\sqrt{PQ}} \cdot \frac{\sqrt{Q}}{2\sqrt{1+P}} .$$

Hence

$$\mathrm{Var}(Y) = \frac{Q}{4+4P}\,\mathrm{Var}\Big[\frac{\sqrt{n}\,(X-P)}{\sqrt{PQ}}\Big] \sim \frac{Q}{4+4P} .$$

This is also the variance of the limiting distribution of Y.

In order to establish the two properties

$\lim_{n\to\infty} [E(T) - \sqrt{n}\,\sinh^{-1}\sqrt{P}\,] = 0$ and $\lim_{n\to\infty} \sigma_T^2 = \frac{Q}{4+4P}$, we follow the agreement of the inverse sine case (see 2.2.4).

We have here

$$q_n(w) = \Psi_n(2w\sqrt{\frac{P(1+P)}{n}} + P) \cdot 2\sqrt{\frac{P(1+P)}{n}}$$

$$= [1 + \frac{2w}{\sqrt{nP}}(\frac{1+2P}{\sqrt{1+P}}) + \frac{4w^2}{n}]^{-\frac{1}{2}} \rightarrow 1 \quad \text{as } n \rightarrow \infty$$

and

$$W = f[Y[\Psi_n(\mu_n)]^{-1} + \mu_n] - f(\mu_n)$$

$$= f[2Y\sqrt{\frac{P(1+P)}{n}} + P] - \sqrt{n}\,\sinh^{-1}P$$

$$= \sqrt{n}\,\sinh^{-1}\sqrt{P+2Y\sqrt{\frac{P(1+P)}{n}}} - \sqrt{n}\,\sinh^{-1}\sqrt{P} \quad ,$$

being the same expression as in the inverse sine case after replacing Q by 1+P. Hence by Schlomilch's mean value theorem we have:

$$W = 2Y[\frac{1-\theta}{(1+2\theta Y\sqrt{\frac{1+P}{nP}})(1-2\theta Y\sqrt{\frac{P}{n(1+P)}}}]^{\frac{1}{2}}$$

$$0 < \theta < 1, \; -\tfrac{1}{2}\sqrt{\frac{nP}{1+P}} < Y < \tfrac{1}{2}\sqrt{\frac{n(P+1)}{P}} \quad .$$

For condition (iii) to be established, the same kind of argument used in the inverse sine case holds.

Since the limiting variance of this transformation involves the parameter P, it is not a satisfactory solution to the stabilization problem, but it is perhaps of some interest that the distribution becomes asymptotically normal.

If P is allowed to vary with n in such a way that $\lim_{n\to\infty} nP = \infty$, it is known that the reduced distribution of X will still tend to normality. If we suppose that $\lim_{n\to\infty} P = 0$ but $\lim_{n\to\infty} nP = \infty$, we find by theorem (2.2.4) that the limiting distribution of

$$Y = \frac{\sqrt{n}\ (X-P)}{\sqrt{PQ}} \cdot \tfrac{1}{2}\sqrt{\frac{1-P}{1+P}}$$

will be normal with mean zero and variance $\frac{1}{4}$. It is easily verified that the conditions (ii) and (iii) of theorem (2.2.3) are still satisfied, so we find the limiting distribution of

$$\sqrt{n}\ \sinh^{-1}\sqrt{X} - \sqrt{n}\ \sinh^{-1}\sqrt{P}$$

is normal with mean zero and variance $\frac{1}{4}$.

(b) If

$$T = f(X) = \begin{cases} \sqrt{n}\ \log X & X > 0 \\ 0 & X \leq 0 \end{cases},$$

then the distribution of $T - \sqrt{n}\ \log P$ tends as $n \to \infty$ to a normal distribution which has mean zero and variance $\frac{Q}{P}$.

Setting here $\Psi_n(x) = f'(x) = \frac{\sqrt{n}}{x}$ when $X > 0$ and $\Psi_n(x) = 0$ otherwise, we put

$$Y = (X-P) \cdot \frac{\sqrt{n}}{P} = \frac{\sqrt{n}\ (X-P)}{\sqrt{PQ}} \cdot \sqrt{\frac{Q}{P}} \quad .$$

Then in analogy with the preceding cases,

$$\operatorname{Var}(Y) = \frac{Q}{P} \operatorname{Var}\left(\frac{\sqrt{n}\ (X-P)}{\sqrt{PQ}}\right) \sim \frac{Q}{P}$$

$$q_n(w) = \frac{P}{\sqrt{n}} \Psi_n[w \cdot \frac{P}{\sqrt{n}} + P] = \frac{P}{\sqrt{n}} \cdot \frac{\sqrt{n}}{w\frac{P}{\sqrt{n}} + P}$$

$$= 1 - \frac{wP}{wP+P\sqrt{n}} \to 1 \quad \text{as } n \to \infty \quad .$$

Thus condition (ii) is satisfied but condition (iii) is not satisfied because

$$W = f[Y\frac{P}{\sqrt{n}} + P] - \sqrt{n} \log P$$

$$= \sqrt{n} \log(\frac{P}{\sqrt{n}} Y + P) - \sqrt{n} \log P$$

$$= \sqrt{n} \log(1 + \frac{Y}{\sqrt{n}}) \quad ,$$

and we are faced with the problem of proving directly that the improper integral $\int_{-\sqrt{n}}^{\infty} n \log(1 + \frac{Y}{\sqrt{n}})^2 \, d\, F_n(y)$ converges uniformly.

2.4 The logarithmic transformation:

2.4.1 Introduction: We shall suppose throughout this section that X is a variate whose mean is μ_n and standard deviation $\sigma = k_n(\mu_n + \alpha)$, where α is an arbitrary constant, $k_n > 0$, and $\lim_{n\to\infty} k_n$ exists and is finite. If k_n is constant for all n, say $k_n = k > 0$, and if we use the heuristic argument that $\sigma_T = f'(\mu_n)\sigma(\mu_n) = k_n$

or

$$f'(\mu_n) = \frac{c}{\sigma(\mu_n)} = \frac{k_n}{k_n(\mu_n+\alpha)} \quad ,$$

then

$$f(\mu_n) = \int \frac{du}{u+\alpha} = \log(u+\alpha) \quad .$$

Hence we find the transformation $T = \log(X+\alpha)$, $X > -\alpha$. It is the purpose of this section to study the asymptotic properties of this transformation.

The theory of such a transformation differs in certain important respects from that of the transformations considered in section (2.2.2). For one thing our starting point in the study of each transformation was the fact that although $P(X < 0) = 0$, nevertheless the reduced distribution of X tended to normality as $n \to \infty$. But in the present case, if X is a variate such that $P(X \leqq -\alpha) = 0$, then the corresponding reduced variate $Y = (X-\mu_n)/[k_n(\mu_n+\alpha)]$ has a d.f. $F_n(y)$ such that,

$$Pr(X \leqq -\alpha) = 0 \quad ,$$

or

$$\Pr\left[\frac{X-\mu_n}{k_n(\mu_n+\alpha)} \leqq \frac{-\alpha-\mu_n}{k_n(\mu_n+\alpha)}\right] = \Pr(Y \leqq -\frac{1}{k_n}) = 0$$

or
$$F_n(-\frac{1}{k_n}) = 0 \quad .$$

Thus if $\lim_{n\to\infty} k_n = k > 0$, the limiting distribution of Y, if it exists, must have a distribution finction F(y) such that $F(-\frac{1}{k}) = 0$. Therefore the limiting distribution of Y can never be normal if $k > 0$.

Moreover (in contrast to the situation in theorem 2.2.2) if the reduced variate Y does have a limiting distribution, the variate

$$W = \frac{1}{k_n}\log(X+\alpha) - \frac{1}{k_n}\log(\mu_n+\alpha)$$
$$= \int_{\mu_n}^{X} \frac{1}{k_n(u+\alpha)}\,du \quad , \qquad X > -\alpha \quad , \tag{97}$$

may have a limiting distribution which is not the same as that of Y. More specifically we have the following result:

2.4.2 <u>Theorem</u>: Let $P(X \leqq -\alpha) = 0$. Let $\lim_{n\to\infty} k_n = k \geqq 0$, and let $F_n(y)$ be the d.f. of the reduced variate $Y = \frac{X-\mu_n}{k_n(\mu_n+\alpha)}$, and let $H_n(w)$ be the d.f. of the variate W given by (97). If a continuous d.f. F(y) exists such that $\lim_{n\to\infty} F_n(y) = F(y)$

for all y, then

$$\lim_{n\to\infty} H_n(w) = \begin{cases} F[\frac{e^{kw}-1}{k}] & k > 0 \\ F(w) & k = 0 \end{cases} .$$

In fact

$$\begin{aligned} H_n(w) &= Pr(W < w) \\ &= Pr(\frac{1}{k_n} \log \frac{X+\alpha}{\mu_n+\alpha} < w) = Pr(\frac{X+\alpha}{\mu_n+\alpha} < e^{k_n w}) \\ &= Pr(\frac{X+\alpha}{\mu_n+\alpha} - 1 < e^{k_n w} - 1) = Pr(\frac{X-\mu_n}{\mu_n+\alpha} < e^{k_n w} - 1) \\ &= Pr(\frac{X-\mu_n}{k_n(\mu_n+\alpha)} < \frac{e^{k_n w}-1}{k_n}) = Pr(Y < \frac{e^{k_n w}-1}{k_n}) \\ &= F_n[\frac{e^{k_n w}-1}{k_n}] \qquad k_n > 0, \quad -\infty < w < +\infty . \end{aligned}$$

The range of Y is $-\frac{1}{k_n} < Y < \infty$.

Hence $$\lim_{n\to\infty} H_n(w) = F[\frac{e^{kw}-1}{k}] , \tag{98}$$

since $k_n \to k$, $F_n \to F$.

In the case $\lim_{n\to\infty} k_n = 0$, we have $\lim_{n\to\infty} H_n(w) = F(\frac{0}{0})$. Since the numerator and denominator of the function $\frac{e^{kw}-1}{k}$ are continuous in k, we can apply l'Hôpital's rule to get

$$\lim_{k\to 0} (\frac{e^{kw}-1}{k}) = \lim_{k\to 0} (\frac{w\ e^{kw}}{1}) = w .$$

Hence

$$\lim_{n\to\infty} H_n(w) = F(w)$$

when $k = 0$.

2.4.3 <u>Theorem</u>: Under the hypothesis of theorem (2.4.2) and under the additional conditions that the improper integral $\int_{-\infty}^{0} w^2 dH_n(w)$ or $\left(\int_{-\frac{1}{k_n}}^{0} k_n^{-2}[\log(1+k_n y)]^2 dF_n(y)\right)$ converges uniformly in n and that $\int_{-\infty}^{+\infty} y^2 dF(y) = 1 = E(y^2)$, the following relations hold:

$$\lim_{n\to\infty} E(W) = \begin{cases} \displaystyle\int_{-\frac{1}{k}}^{\infty} \frac{1}{k} \log(1+ky)\, dF(y) & k > 0 \\ 0 & k = 0 \end{cases} , \qquad (99)$$

$$\lim_{n\to\infty} E(W^2) = \begin{cases} \displaystyle\int_{-\frac{1}{k}}^{\infty} \frac{1}{k^2}[\log(1+ky)]^2 dF(y) & k > 0 \\ 1 & k = 0 \end{cases} . \qquad (100)$$

By theorem (2.4.2) we know that

$$E(W) = \int_{-\infty}^{+\infty} w\, d\, H_n(w) = \int_{-\infty}^{+\infty} w\, d\, F_n\left(\frac{e^{k_n w} - 1}{k_n}\right) .$$

Making the change of variable $\frac{e^{k_n w} - 1}{k_n} = y$, we have

$$w = \frac{1}{k_n} \log(1+k_n y) \quad . \tag{101}$$

The w ranges from $-\infty$ to $+\infty$ as y ranges from $-\frac{1}{k_n}$ to ∞.

Thus we have

$$E(W) = \int_{-\frac{1}{k_n}}^{\infty} \frac{1}{k_n} \log(1+k_n y)\, dF_n(y) \quad .$$

As $n \to \infty$, $F_n(y) \to F(y)$, $k_n \to k \geqq 0$ so

$$\lim_{n\to\infty} E(W) = \begin{cases} \displaystyle\int_{-\frac{1}{k}}^{\infty} \frac{1}{k} \log(1+ky)\, dF(y) & k > 0 \\ 0 & k = 0 \end{cases} \quad .$$

Similarly for $\lim_{n\to\infty} E(W^2)$, but when $k = 0$ the limit $\frac{1}{k^2}[\log(1+ky)]^2$ is of the form $\frac{0}{0}$. Applying l'Hôpital's rule we get

$$\lim_{k\to 0} \frac{[\log(1+ky)]^2}{k^2} = \lim_{k\to 0} \frac{2[\log(1+ky)] \cdot \frac{y}{1+ky}}{2k}$$

$$= \lim_{k\to 0} \frac{y^2/(1+ky)}{yk+(1+ky)} = y^2 \quad ,$$

hence in the case $k = 0$

$$\lim_{n\to\infty} E(W^2) = \int_{-\infty}^{+\infty} y^2\, dF(y) = 1 \quad .$$

Thus

$$\lim_{n\to\infty} E(W^2) = \begin{cases} \displaystyle\int_{-\frac{1}{k}}^{\infty} \frac{1}{k^2}[\log(1+ky)]^2\, d\,F(y) & k > 0 \\ 1 & k = 0 \end{cases} .$$

2.4.4 Remarks: The variance σ_T^2 of the variate $T = \log(X+\alpha)$ is given by the equation $\sigma_T^2 = k_n^2\{E(W^2) - [E(W)]^2\}$. Thus if $F(y)$ is independent of any unknown parameter θ, and if k is positive and is presumed to be constant, then the transformation $T = \log(X+\alpha)$ is seen to yield an asymptotic stabilization of the variance under the conditions of theorem (2.4.3).

Theorem (2.4.2) raises the following question: Just what limiting distribution must Y have if $k > 0$ in order that the distribution of W tends to normality? To answer this we shall note the following non-asymptotic result.

2.4.5 Theorem: A necessary and sufficient condition that X have a continuous distribution with density function:

$$\varphi(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi\log(1+k^2)}} \cdot \dfrac{1}{x+\alpha} \exp\left[\dfrac{-\left(\log\dfrac{(x+\alpha)\sqrt{k^2+1}}{\mu+\alpha}\right)^2}{2\log(k^2+1)}\right], & x > -\alpha \\ 0 & x \leq -\alpha \end{cases} \tag{102}$$

(for which $\sigma_X = k(\mu+\alpha)$), is that the variate $T = \log(X+\alpha)$

have a normal distribution with mean = $\log(\mu+\alpha) - \log\sqrt{k^2+1}$ and variance = $\log(1+k^2)$. The proof needs a routine change of variable in the normal distribution of T. We know that

$$dF(T) = \frac{1}{\sqrt{2\pi\log(k^2+1)}} \exp\left[\frac{-(T-\log(\mu+\alpha)+\log\sqrt{k^2+1})^2}{2\log(k^2+1)}\right] dT \quad .$$

Making the change of variable $T = \log(x+\alpha)$, we get $\varphi(x)$ as it is represented in (102).

If X is distributed according to (102), the density function $F'(y)$ of the corresponding reduced variate $Y = (x-\mu)/[k(\mu+\alpha)]$ is:

$$F'(y) = \begin{cases} \dfrac{k}{\sqrt{2\pi\log(1+k^2)}} \dfrac{1}{1+ky} \exp\left[\dfrac{-\{\log[(1+ky)\sqrt{1+k^2}]\}^2}{2\log(1+k^2)}\right] & y > -\dfrac{1}{k} \\ 0 & y \leq -\dfrac{1}{k} \end{cases} \quad (103)$$

The distribution function of the variate $W = k^{-1}[\log(X+\alpha) - \log(\mu+\alpha)]$ is $F[\frac{e^{kw}-1}{k}]$ as we have seen in theorem (2.4.2), since $\log(X+\alpha)$ is normal with mean = $\log(\mu+\alpha) - \log\sqrt{1+k^2}$ and variance = $\log(1+k^2)$. Thus W is normal with mean = $-k^{-1}\log\sqrt{k^2+1}$ and variance = $k^{-2}\log(k^2+1)$. The quantity $-k^{-1}\log\sqrt{k^2+1}$ is the value of the integral in (99), since the integral becomes, after substituting $dF(y) = F'(y)dy$ and making the change of variable $\log(1+ky)=t$,

$$\frac{1}{\sqrt{2\pi\log(k^2+1)}}\int_{-\infty}^{+\infty}\frac{t}{e^t}\exp\left[\frac{-(t+\log\sqrt{k^2+1})^2}{2\log(k^2+1)}\right]\frac{1}{k}\,e^t dt = \frac{1}{k}E(t) \quad ,$$

where t is $N(-\log\sqrt{k^2+1},\ \log(k^2+1))$. Hence the integral equals $(-k^{-1}\log\sqrt{k^2+1})$ as indicated above.

Similarly the value of the integral in (100) becomes:

$$\frac{1}{\sqrt{2\pi\log(k^2+1)}}\int_{-\infty}^{\infty}\frac{t^2}{e^t}\exp\left[\frac{-(t+\log\sqrt{k^2+1})^2}{2\log(k^2+1)}\right]\frac{1}{k}\,e^t dt$$

$$= \frac{1}{k^2}E(t^2) = k^{-2}[\log(k^2+1)+(\log\sqrt{k^2+1})^2] \quad .$$

If now the distribution of X depends on a parameter n in such a way that as $n \to \infty$, the distribution of the corresponding reduced variate $Y = (X-\mu_n)/k_n(\mu_n+\alpha)$ tends to the distribution given by (103), it follows from the above discussion and from theorem (2.4.2) that the variate $W = \frac{1}{k_n}\log(X+\alpha) - \frac{1}{k_n}\log(\mu_n+\alpha)$ has a normal limiting distribution. Furthermore, under the uniform convergence condition of theorem (2.4.3), it follows that σ_T^2 tends to the value $\log(k^2+1)$ where $T = \log(X+\alpha)$.

2.4.6 <u>Discussion</u>: The above theorems and remarks provide a mathematical basis for the use of the logarithmic transformation.

When it appears from a reasonably large number of observations on a variate (which is essentially bounded from below) that the standard deviation of the variate is proportional to the mean, then a possible specification for the variate is a distribution of the form (102). If such specification is sound then the variate $T = \log(X+\alpha)$, where $-\alpha$ is any number less than the lower bound of X, will be exactly or approximately normally distributed with a variance independent of the value of μ. But since (102) is only one of an infinity of different types of distribution in which the mean and standard deviation are proportional, the user of a log#arithmetic transformation in the analysis of variance should always apply tests to the T-values for departures from normality. From the point of view of specification, the situation here would seem to be less reassuring than in the cases considered in the angular or square root transformation. While it is true, for instance, that the Poisson distribution is only one of many types of distributions in which the variance and mean are equal, nevertheless the specification of a Poisson distribution can generally be preceded by a fairly strong reasoning which can be deduced from the nature of the data. This would not seem to be the case in the specification of (102). Theorems (2.4.2) and (2.4.3) furnish some grounds for supposing that the logarithmic transformation may possibly be more successful in stabilizing the variance than in normalizing the data.

2.5 On the distribution of a variate whose logarithm is normally distributed:

2.5.1 Moments of the distribution: Finney (1941) examined the moments of such a distribution and their estimates. Let y be a variate whose distribution is such that $x = \log_e y$ is normally distributed (i.e. the distribution $\varphi(x)$ mentioned in the previous section, equation (102)) with mean ξ and variance σ^2. The probability density of x is then:

$$f(x) = \frac{1}{\sqrt{2\pi}\ \sigma} e^{-\frac{1}{2\sigma^2}(x-\xi)^2} \quad . \tag{104}$$

We know that $E(e^{itx}) = e^{it\xi+\frac{1}{2}(it)^2\sigma^2}$ is the characteristic function of $N(\xi, \sigma^2)$. Replacing (it) by r we have $E(e^{rx}) = e^{r\xi+\frac{1}{2}r^2\sigma^2}$. Since $y = e^x$, it follows that,

$$E(y^r) = e^{r\xi+\frac{1}{2}r^2\sigma^2} \quad , \tag{105}$$

which defines the cumulants (and thus the moments). In particular

$$\mu = e^{\xi+\frac{1}{2}\sigma^2}$$
$$\mu_2(y) = e^{2\xi+\sigma^2}(e^{\sigma^2}-1) \quad . \tag{106}$$

For convenience let us write

$$\tau = e^{\sigma^2} \tag{107}$$

so that

$$\mu_2(y) = \mu^2(\tau-1) \quad . \tag{108}$$

The coefficient of variation is therefore $(\tau-1)^{\frac{1}{2}}$, which takes the value unity when $\sigma^2 = .693$, higher values than this can undoubtedly occur, but the chief practical interest probably concerns populations of less variability. The measures of departure from normality are

$$\gamma_1 = (\tau-1)^{\frac{1}{2}}(\tau+2) \quad ,$$

$$\gamma_2 = (\tau-1)(\tau^3+3\tau^2+6\tau+6) \quad .$$

2.5.2 Estimation of the moments: Suppose that a sample of n individuals is taken from the population. It is known that sufficient statistics for the estimation of the parameters of the transformed distribution are

$$\bar{x} = \frac{\Sigma x}{n} \quad , \tag{109}$$

$$s^2 = \frac{1}{n-1}\,\Sigma(x-\bar{x})^2 \quad . \tag{110}$$

We know that $(n-1)\frac{s^2}{\sigma^2}$ is distributed as χ^2 with n-1 degrees of freedom and so

$$E(s^{2p}e^{as^2}) = \frac{1}{\Gamma(\frac{n-1}{2})}\left(\frac{n-1}{2\sigma^2}\right)^{\frac{1}{2}(n-1)}\int_0^{\infty}(s^2)^{\frac{1}{2}(n-1)+p-1}e^{-(\frac{n-1}{2\sigma^2}-a)s^2}\,d(s^2)$$

$$= \frac{\Gamma(\frac{n-1}{2}+p)}{\Gamma(\frac{n-1}{2})} \cdot \left(\frac{2\sigma^2}{n-1}\right)^p\left(1-\frac{2a\sigma^2}{n-1}\right)^{-\frac{1}{2}(n-1)-p} , \qquad (111)$$

and in particular $(a = 0)$,

$$E(s^{2p}) = \frac{\Gamma(\frac{n-1}{2}+p)}{\Gamma(\frac{n-1}{2})} \cdot \left(\frac{2\sigma^2}{n-1}\right)^p$$

$$= \frac{(\frac{n-3}{2}+p)!}{(\frac{n-3}{2})!} \cdot \left(\frac{2\sigma^2}{n-1}\right)^p$$

$$= \frac{(n+1)(n+3)\ldots(n+2p-3)}{(n-1)^{p-1}}\sigma^{2p} . \qquad (112)$$

If the infinite series $g(t)$ is defined by

$$g(t) = 1 + \frac{n-1}{n}t + \frac{(n-1)^3}{n^2 2!}\frac{t^2}{n+1} + \frac{(n-1)^5}{n^3 3!}\frac{t^3}{(n+1)(n+3)} + \ldots \qquad (113)$$

then

$$E[g(\tfrac{1}{2}r^2s^2)] = E[1+\frac{n-1}{2n}r^2s^2+\frac{(n-1)^3r^4s^4}{n^2 2!4(n+1)}+\frac{(n-1)^5r^6s^6}{n^3 3!8(n+1)(n+3)}+\ldots]$$

$$= 1+\frac{n-1}{n2!}r^2(n-1)\sigma^2+\frac{(n-1)^3r^4\sigma^4(n+1)}{n^2 2!4(n+1)(n-1)}+\frac{(n-1)^5r^6\sigma^6(n+1)(n+3)}{n^3 3!8(n+1)(n-1)^2}+\ldots$$

$$= 1+\frac{(n-1)^2}{n2!}r^2\sigma^2+\frac{(n-1)^2}{n^2 2!4}r^4\sigma^4+\frac{(n-1)^3}{n^3 3!8}r^6\sigma^6+\ldots$$

$$= e^{\frac{n-1}{2n}r^2\sigma^2} . \qquad (114)$$

Also, since the sampling distribution of $\bar{x}$ is normal $N(\xi, \sigma^2/n)$, we have

$$E(e^{r\bar{x}}) = e^{r\xi+\frac{r^2\sigma^2}{2n}} \quad . \tag{115}$$

The distribution of $\bar{x}$ and s^2 are independent, and therefore

$$E[e^{r\bar{x}}\, g(\tfrac{1}{2}r^2 s^2)] = e^{r\xi+\frac{r^2\sigma^2}{2n}+\frac{n-1}{2n}\, r^2\sigma^2}$$

$$= e^{r\xi+\frac{1}{2}r^2\sigma^2} = E(y^r) \quad . \tag{116}$$

Since the statistics are sufficient, any function of them alone must be an efficient estimate of its expected value. Hence efficient estimates of the mean and variance of the y population will be

$$m = e^{\bar{x}}\, g(\tfrac{1}{2}s^2) \tag{117}$$

(putting $r = 1$ in (116)) and

$$v = e^{2\bar{x}}\, [g(2s^2)-g(\tfrac{n-2}{n-1}\, s^2)] \quad . \tag{118}$$

v is chosen so that $E(v) = \mu_2(y)$. In fact we have here,

$$E(v) = E[e^{2\bar{x}}\, g(2s^2)] - E[e^{2\bar{x}}\, g(\tfrac{n-2}{n-1}\, s^2)]$$

$$= E[e^{2\bar{x}}\, g(2s^2)] - E(e^{2\bar{x}})\; E[g(\tfrac{n-2}{n-1}\, s^2)]$$

$$= e^{2\xi+2\sigma^2} - e^{2\xi+\frac{2\sigma^2}{n}}\; e^{(1-\frac{2}{n})\sigma^2}$$

$$= e^{2\xi+2\sigma^2} - e^{2\xi+\sigma^2} = \mu_2(y) \quad ,$$

so that v is an unbiased efficient estimate for $\mu_2(y)$.

Unfortunately the series g(t) is not very suitable for computational purposes, as its convergence is slow, except for small values of t. It is possible that occasionally the value of g might be satisfactorily obtained from its representation in series, but the convergence of v will be very much slower. We can, however, develop the series in ascending powers of n^{-1}, and thus obtain expeditiously for moderately large samples, an approximation to the estimate.

2.5.3 Approximation to the efficient estimate: The estimates m and v given above may be expressed in a form more suitable for arithmetical computation if an expansion of g(t) in ascending powers of n^{-1} is first obtained. For this purpose we write,

$$e^{-t}g(t) = 1 - \frac{t(t+1)}{n} + \frac{t^2(3t^2+22t+21)}{6n^2} + \ldots$$

or

$$g(t) = e^{t}[1 - \frac{t(t-1)}{n} + \frac{t^2(3t^2+22t+21)}{6n^2} + \ldots] \quad . \qquad (119)$$

The approximations to the efficient estimates are to the order of n^{-2},

$$m = e^{\bar{x}+\frac{1}{2}s^2}[1 - \frac{s^2(s^2+2)}{4n} + \frac{s^4(3s^4+44s^2+48)}{96n^2}] \qquad (120)$$

and

$$v = e^{\overline{x}^2+s^2}\{e^{s^2}[1 - \frac{2s^2(2s^2+1)}{n} + \frac{2s^4(12s^4+44s^2+21)}{3n^2}]$$

$$- [1 - \frac{s^2(s^2+2)}{n} + \frac{s^4(3s^4+28s^2+42)}{6n^2}]\} \quad . \qquad (121)$$

For values of the sample variance s^2, of the transformed populations which approach 0.69, corresponding to 100 percent standard deviation in the y distribution (i.e., coefficient of variation = 1), possibly $n > 50$ in (120) and $n > 100$ in (121) would be safe limits for practical purposes of estimation.

2.5.4 Efficiency of estimation without transformation: It is of interest to determine the efficiency with which the mean and variance of the distribution of y are determined without using a transformation. From the first and second moments of the sample the estimates are:

$$\overline{y} = \frac{\Sigma y}{n} \quad , \qquad (122)$$

$$D = \frac{\Sigma(y-\overline{y})^2}{n-1} \quad . \qquad (123)$$

The sampling variances of these are

$$V(\overline{y}) = \frac{V(y)}{n} = \frac{\mu^2(\tau-1)}{n} \qquad (124)$$

$$V(D) = \frac{K_4}{n} + \frac{K_2^2}{n-1}$$

$$= \mu^4(\tau-1)^2[(\tau^4+2\tau^3+3\tau^2-4)+\ldots]/n \quad , \qquad (125)$$

where K_2 and K_4 are found from (105).

In order to obtain the variance of the efficient estimate, m, it is necessary to find $E(m^2)$. Now from (111), by the use of Stirling's approximation,

$$E(s^{2p}e^{as^2}) = \sigma^{2p}e^{a\sigma^2}\{1+\frac{a^2\sigma^4+p(2a\sigma^2+p-1)}{n}+\ldots\} \quad . \qquad (126)$$

A more detailed expansion for the case p = 0 gives:

$$E(e^{as^2}) = e^{a\sigma^2}\{1+\frac{a^2\sigma^4}{n}+\frac{a^2\sigma^4(3a^2\sigma^4+8a\sigma^2+6)}{6n^2}+\ldots\} \quad . \qquad (127)$$

It follows from (119) that

$$E\{[g(\tfrac{1}{2}s^2)]^2\}=E\{e^{\frac{1}{2}s^2}[1-\frac{s^2(s^2+2)}{4n}+\frac{s^4(3s^4+44s^2+48)}{96n^2}+\ldots]\}^2$$

$$= E\{e^{s^2}[1-\frac{s^2(s^2+2)}{2n}+\frac{s^4(3s^4+44s^2+84)}{48n^2}+\frac{s^4(s^4+4s^2+4)}{16n^2}+\ldots]\}$$

$$= E\{e^{s^2}[1-\frac{s^2(s^2+2)}{2n}+\frac{s^4(3s^4+28s^2+48)}{24n^2}+\ldots]\}$$

$$= E\{e^{s^2}-\frac{e^{s^2}s^4}{2n}-\frac{s^2e^{s^2}}{n}+\frac{e^{s^2}s^8}{8n^2}+\frac{7e^{s^2}s^6}{6n^2}+\frac{2s^4e^{s^2}}{n^2}+\ldots\} \quad .$$

The equations (126) and (127) give the required expectations for each term, so we get:

$$E\{[g(\tfrac{1}{2}s^2)]^2\} = e^{\sigma^2}\{1 + \frac{\sigma^4-2\sigma^2}{2n} + \frac{\sigma^8-4\sigma^6}{8n^2} + \ldots\}$$

and

$$E(m^2)=E(e^{2\overline{x}})E\{[g(\tfrac{1}{2}s^2)]^2\}=e^{2\xi+\frac{2\sigma^2}{n}}e^{\sigma^2}\{1+\frac{\sigma^4-2\sigma^2}{2n}+\frac{\sigma^8-4\sigma^6}{8n^2}+\ldots\}$$

$$= e^{2\xi+\sigma^2}[1 + \frac{2\sigma^2}{n} + \frac{4\sigma^4}{2n^2} + \ldots][1 + \frac{\sigma^4-2\sigma^2}{2n} + \frac{\sigma^8-4\sigma^6}{8n^2} + \ldots]$$

$$= e^{2\xi+\sigma^2}[1 + \frac{\sigma^4+2\sigma^2}{2n} + \frac{\sigma^8+4\sigma^6}{8n^2} + \ldots] \quad .$$

Now

$$\mathrm{Var}(m)=E(m^2)-(E(m))^2=e^{2\xi+\sigma^2}[1+\frac{\sigma^4+2\sigma^2}{2n}+\frac{\sigma^8+4\sigma^6}{n^2}+\ldots]-e^{2\xi+\sigma^2}$$

$$= e^{2\xi+\sigma^2}[\sigma^2 + \frac{\sigma^4}{2} + \frac{1}{2n}(\sigma^6 + \frac{\sigma^8}{4})]/n \quad . \qquad (128)$$

The efficiency of the estimate $\overline{y}$ is thus:

$$\text{Efficiency } (\overline{y})=\frac{V(m)}{V(\overline{y})}=\frac{e^{2\xi+\sigma^2}\{\sigma^2+\frac{\sigma^4}{2}+\frac{1}{2n}(\sigma^6+\frac{\sigma^8}{4})\}/n}{e^{2\xi+\sigma^2}(\tau-1)/n}$$

$$= \{\sigma^2+\frac{\sigma^4}{2}+\frac{1}{2n}(\sigma^6+\frac{\sigma^8}{4})\}/\tau-1 \quad . \qquad (129)$$

In large samples, and for small values of σ^2, the efficiency is almost 100 percent, and never falls below 93 percent when σ^2 is less than 0.7; even for $\sigma^2 = 2.5$ the efficiency is still 50 percent. The efficiency increases slightly with smaller values of n.

By a similar process, taking the expectation of v^2 using the formulae (126), (127) and remembering that $\overline{x}$ and s^2

are independent, we get,

$$E(v^2)=e^{4\xi}[e^{4\sigma^2}\{1+\frac{8\sigma^4+4\sigma^2}{n}\}-2e^{3\sigma^2}\{1+\frac{4\sigma^4+4\sigma^2}{n}\}+e^{2\sigma^2}\{1+\frac{2\sigma^4+4\sigma^2}{n}\}]$$

and thus,

$$V(v) = E(v^2) - [E(v)]^2 = E(v^2) - \mu^4(\tau-1)^2$$

$$= e^{4\xi} + 2\sigma^2\{4\sigma^2(\tau-1)^2+2\sigma^4(2\tau-1)^2\}/n \quad . \qquad (130)$$

It follows that,

$$\text{Efficiency (D)} = \frac{4\sigma^2(\tau-1)^2+2\sigma^4(2\tau-1)^2}{(\tau-1)^2(\tau^4-2\tau^3+3\tau^2-4)} \quad . \qquad (131)$$

This quantity also approaches 100 percent for small values of σ^2, but falls rapidly with increasing σ^2. Thus for $\sigma^2 = 0.1$ the efficiency is only 79 percent and for $\sigma^2 = .69$ it is 28 percent. The form taken by this function when terms in n^{-1} are included has not been fully investigated, but such an examination could scarcely alter the general conclusion that the use of D as an estimate of the variance of the y distribution is inefficient except for the smallest values of σ^2.

The large sample efficiencies of the untransformed estimates $\overline{y}$ and D are shown in Fig. VII*for values of σ^2 up to 2.0.

---

*See page 177.

2.6 <u>The statistical analysis of variance-heterogeneity and the logarithmic transformation</u>: While a useful approximate null test of significance of the heterogeneity of variances is available (Bartlett test), it is often necessary in more detailed investigations of variance heterogeneity to apply the powerful technique of "analysis of variance" to the data, when suitably transformed. For an estimate s^2 of a variance σ^2 based on n d.f., the distribution of $\frac{ns^2}{\sigma^2}$ is well known to be a χ^2 distribution with n d.f, if the estimate s^2 has been obtained from a normal sample. The density function of $\frac{ns^2}{\sigma^2}$ is $\frac{1}{\Gamma(\frac{n}{2})}(\frac{ns^2}{2\sigma^2})^{\frac{n}{2}-1} e^{-\frac{ns^2}{2\sigma^2}} d(\frac{ns^2}{2\sigma^2})$. It follows that the distribution of $\ell n\ s^2 - \ell n\ \sigma^2$ is entirely independent of σ^2, and hence that that of $\ell n\ s^2$ only depends on σ through the term $\ell n\ \sigma^2$ in its mean value. The variate $\ell n\ s^2$ is thus a convenient variate to consider. We know that $E(s^2) = \sigma^2$, $V(s^2) = \frac{2\sigma^4}{n}$ or $\sqrt{V(s^2)} = \sqrt{\frac{2}{n}}\, E(s^2)$, i.e., the standard deviation is proportional to the mean. Hence the logarithmic transformation is the appropriate one for our purpose. Bartlett and Kendall (1946) studied the properties of this transformation and the following is a summary of their work:

For assessing the normality of the transformed variate $\ell n\ s^2$, let us take its characteristic function

$$M(t) = \frac{1}{\Gamma(\frac{n}{2})} \int_0^\infty (s^2)^{it} (\frac{ns^2}{2\sigma^2})^{\frac{n}{2}-1} e^{-\frac{ns^2}{2\sigma^2}} d(\frac{ns^2}{2\sigma^2})$$

$$= (\frac{2\sigma^2}{n})^{it} \frac{1}{\Gamma(\frac{n}{2})} \int_0^\infty (\frac{ns^2}{2\sigma^2})^{\frac{n}{2}+it-1} e^{-\frac{ns^2}{2\sigma^2}} d(\frac{ns^2}{2\sigma^2})$$

$$= (\frac{\sigma^2}{n/2})^{it} \Gamma(\frac{n}{2} + it)/\Gamma(\frac{n}{2}) \quad , \qquad (132)$$

where the cumulant function $K(t) = \ell n\ M(t)$ is

$$K(t) = it(\ell n\ \sigma^2 - \ell n\ \tfrac{1}{2}n) + \ell n\ \Gamma(\frac{n}{2} + it) - \ell n\ \Gamma(\frac{n}{2}) \quad .$$

Expanding the function $\ell n\ \Gamma(\frac{n}{2} + it)$ in a Maclaurin series around $t = 0$, we get

$$\ell n\ \Gamma(\frac{n}{2} + it) = \ell n\ \Gamma(\frac{n}{2}) + \frac{it}{1!}\Psi(\frac{n}{2}) + \frac{(it)^2}{2!}\Psi'(\frac{n}{2}) \ \ldots \quad , \qquad (133)$$

where $\Psi(x) = \frac{d(\ell n\ \Gamma(x))}{dx}$, and $\Psi^{(r)}(x)$ is the $(r+1)^{th}$ derivative of $\ell n\ \Gamma(x)$. Hence

$$K_1 = (\ell n\ \sigma^2 - \ell n\ \frac{n}{2}) + \Psi(\frac{n}{2}) \quad , \qquad (134)$$

$$K_{r+1} = \Psi^{(r)}(\frac{n}{2}) \qquad (r > 0) \quad .$$

From these results the values of K_1, K_2, $\gamma_1 = \frac{K_3}{(K_2)^{3/2}}$ and $\gamma_2 = \frac{K_4}{K_2^{\ 2}}$ may be computed and are given for reference in Table VII up to $n = 20$. For larger values of n, it is

Table VII. Constants for the distribution of $\ell n\ s^2$.

n	$K_1 - \ell n\sigma^2$	$-(\frac{1}{n}+\frac{1}{3n^2})$	$\frac{1}{K_2}$	$\frac{2}{(n-1)K_2}$	ν_1	ν_2	$E=(\frac{2}{nk_2})x100$
1	-1.27036	-1.33333	0.20264	--	-1.535	+4.000	40.53
2	.57721	.58333	0.60793	1.2159	1.140	2.400	60.79
3	.36898	.37037	1.0697	1.0697	.917	1.613	71.30
4	.27036	.27083	1.5505	1.0337	.780	1.188	77.53
5	.21313	.21333	2.0393	1.0197	.688	0.931	81.57
6	.17583	.17593	2.5321	1.0128	.621	0.763	84.40
7	.14961	.14966	3.0270	1.0090	.570	0.644	86.49
8	.13018	.13021	3.5233	1.0067	.529	0.557	88.08
9	.11521	.11523	4.0205	1.0051	.496	0.490	89.34
10	.10332	.10333	4.5183	1.0041	.469	0.437	90.37
11	.09365	.09366	5.0165	1.0033	.445	0.395	91.21
12	.08564	.08565	5.5150	1.0027	.425	.360	91.92
13	.07889	.07890	6.0138	1.0023	.407	.330	92.52
14	.07313	.07313	6.5128	1.0020	.391	.305	93.04
15	.06815	.06815	6.0119	1.0017	.377	.284	93.49
16	.06380	.06380	7.5111	1.0015	.364	.265	93.89
17	.05998	.05998	8.0104	1.0013	.353	.249	94.24
18	.05658	.05658	8.5098	1.0012	.342	.234	94.55
19	.05355	.05356	9.0092	1.0010	.333	.221	94.83
20	-.05083	-.05083	9.5088	1.0009	-.324	+.210	95.09

sufficient to take

$$
\begin{aligned}
K_1 &\sim -\left(\frac{1}{n} + \frac{1}{3n^2}\right) + \ell n\ \sigma^2 \quad , \\
K_2 &\sim \frac{2}{n-1} \\
\gamma_1 &\sim \sqrt{\frac{2}{n-1}} \\
\gamma_2 &\sim \frac{4}{n-1}
\end{aligned}
\qquad (135)
$$

The values of γ_1 and γ_2 are also plotted in Fig. VIII,* and the ratio of $\frac{2}{n-1}$ to K_2 in Fig. IX.* For the efficiency of the mean on the transformed scale, it is noted that in the case of complete homogeneity the information in the mean value of a set of statistics $\ell n\ s_r^2$ tends, as the number in the set becomes large, to be proportional to $\frac{1}{K_2}$, whereas the information in the sufficient statistics s^2 is proportional to $\frac{n}{2}$, $(\mathrm{Var}(s^2) = \frac{2\sigma 4}{n})$. A measure of the efficiency is thus $E = \frac{2}{nK_2}$ and this is also given in Table VII and plotted in Fig. X.*

While no hard-and-fast rule can be laid down, the above results suggest that the transformation may safely be used for $n = 10$ and over, more tentatively from $n = 5$ to 9, and not at all below $n = 5$. In the first case E is over 90 percent, and in the second case between 80 percent and 90

---

*See pages 177-178.

percent, while in the last case E falls below 80 percent. The values of γ_1 and γ_2 indicate however that the approach to normality is rather slow.

2.7 Basic theorem and application:

2.7.1 Theorem: Eisenhart (1947) gave a basic theorem, although it is less appropriate from the mathematical point of view than the approach of Curtiss. However, it is more practical and easy to apply. The theorem is as follows.

Let θ $(\theta_1 \leqq \theta \leqq \theta_2)$, be a parameter of a statistical universe (population) and let T be an unbiased estimator of θ, $[T_1(\theta) \leqq T \leqq T_2(\theta)]$, based on n observations from the universe with

$$\begin{aligned} \mu_1(T) &\equiv \theta \\ \mu_2(T) &\equiv \frac{g(\theta)}{n} \\ \mu_3(T) &\equiv 0(\frac{1}{n}) \quad \text{as } n \to \infty \\ \mu_4(T) &\equiv 0(\frac{1}{n}) \quad \text{as } n \to \infty \quad . \end{aligned} \tag{136}$$

If the function

$$f(T) = \int \frac{dt}{\sqrt{g(T)}} \tag{137}$$

and its first two derivatives exist for all admissible values of T, except perhaps for a set of measure zero, and if the mean value of its second derivative and of the square of the second derivative are finite for $\theta_1 < \theta < \theta_2$ and every value of n, that is if

$$E[f^{(2)}(T)] = O(1)$$
$$E[f^{(2)}(T)]^2 = O(1) \tag{138}$$

as $n \to \infty$ for $\theta_1 < \theta < \theta_2$ and $T_1(\theta) \leqq T \leqq T_2(\theta)$, then

$$V[f(T)] = \frac{1}{n} + O(\frac{1}{n}) \quad \text{as } n \to \infty \quad . \tag{139}$$

<u>Proof</u>: Let f(T) be a function of T such that its second derivative $f^{(2)}(T)$ exists for $T_1(\theta) < T < T_2(\theta)$, except perhaps for a set of points of measure zero, where $T_1(\theta) \leqq T \leqq T_2(\theta)$; that is, $p[T_1(\theta) \leqq T \leqq T_2(\theta)] = 1$. Since $E(T) \equiv \theta$ by assumption, it follows that θ lies in the admissible range of values for T. Hence, using Taylor's theorem, f(T) may be expressed as follows:

$$f(T) = f(\theta) + \frac{T-\theta}{1!}f^{(1)}(\theta) + \frac{(T-\theta)^2}{2!}f^{(2)}(t) \quad , \tag{140}$$

where $t = \theta + \xi_T(T-\theta)$, $0 < \xi_T < 1$, and the notation ξ_T is used to emphasize the fact that the value of ξ depends upon the value of T concerned, with $T_1(\theta) \leqq T \leqq T_2(\theta)$.

Taking the expectation of each side of equation (140), we find that

$$E[f(T)] = f(\theta) + \frac{V(T)}{2!} E[f^{(2)}(t)] = f(\theta)+0(1) \quad \text{as } n\to\infty \quad (141)$$

by virtue of assumptions (138). This shows that these assumptions, together with those regarding the existence and differentiability of f(T), are sufficient to ensure that, as n increases without limit, the expected value of the function f(T) converges to $f(\theta)$.

Subtracting the left and right sides of (141) from the left and right sides, respectively, of (140), squaring, and taking expected values, it is found that:

$$V[f(T)] \equiv E\{f(T)-E[f(T)]^2 = [f^{(1)}(\theta)]V(T) + \frac{3E^2[f^{(2)}(t)]V^2(T)}{4}$$

$$+ \frac{E[f^{(2)}(t)]^2\mu_4(T)}{4} + \frac{f^{(1)}(\theta)E[f^{(2)}(t)]\mu_3(T)}{1} . \quad (142)$$

In view of the assumptions (136), this becomes:

$$V[f(T)]=[f^{(1)}(\theta)]^2V(T)+0(\frac{1}{n})=[f^{(1)}(\theta)]^2 \frac{g(\theta)}{n} + 0(\frac{1}{n}) \quad (143)$$

as $n \to \infty$.

Consequently, for the function defined by equation (137), equation (142) reduces to equation (139). This completes the proof.

2.7.2 Applications:

(a) Mean of a sample from a Poisson population:

Let $\theta = \mu$, the parameter (mean) of a Poisson population, and let $T = \bar{x}$, the mean of a random sample of n independent observations from this population. It is well known that

$$\begin{aligned} E(\bar{x}) &= \mu \\ V(\bar{x}) &= \frac{\mu}{n} \\ \mu_3(\bar{x}) &= \frac{\mu}{n^2} \\ \mu_4(\bar{x}) &= \frac{3\mu^2}{n^2} + \frac{\mu}{n^3} \quad . \end{aligned} \tag{144}$$

Clearly here $\mu_3(\bar{x})$ and $\mu_4(\bar{x})$ are both $0(\frac{1}{n})$ as $n \to \infty$, and $g(\mu) = \mu$. Consequently, the theorem implies that if there exists a transformation that will stabilize the variance, it will be

$$f(\bar{x}) = \int \frac{d\bar{x}}{\sqrt{\bar{x}}} = 2\sqrt{\bar{x}} + c \quad . \tag{145}$$

It is convenient to take $c = 0$, and drop the scale factor 2, adopting the transformation $y = \sqrt{\bar{x}}$ (referred to previously). The first and second derivatives of y with respect to $\bar{x}$ are $y^{(1)} = \frac{1}{2\sqrt{\bar{x}}}$, $y^{(2)} = \frac{-1}{4(\bar{x})^{3/2}}$, and approximately

$$E(y^{(1)}) = \frac{1}{2\sqrt{\mu}}$$

$$E(y^{(2)}) = \frac{-1}{4(\mu)^{3/2}}$$

$$[E(y^{(2)})]^2 = \frac{1}{16\mu^3} \quad .$$

Hence equation (142) becomes:

$$V(y) = \frac{1}{4n} + \frac{3}{64\mu n^2} + \frac{1}{64}\left(\frac{3}{\mu n^2} + \frac{1}{\mu^2 n^3}\right) - \frac{1}{8\mu n^2}$$

$$= \frac{1}{4n} - \frac{1}{32\mu n^2} + \frac{1}{64\mu^2 n^3} \quad , \qquad \text{approximately.}$$

Therefore, as $n \to \infty$, for fixed μ,

$$V(y) = \frac{1}{4n} + 0\left(\frac{1}{n}\right) \quad , \tag{146}$$

and as $\mu \to \infty$ for fixed n

$$V(y) = \frac{1}{4n} + 0\left(\frac{1}{\mu}\right) \quad . \tag{147}$$

Consequently if either n or μ is large, $V(y) = \frac{1}{4n}$ approximately.

(b) Estimate of the variance of a normal population based on n degrees of freedom:

Let $\theta = \sigma^2$ be the variance of a normal population and let $T = s^2$ be an estimate of σ^2 such that $\frac{ns^2}{\sigma^2}$ is distributed as χ^2 for n degrees of freedom. Then

$$E(s^2) = \sigma^2$$

$$V(s^2) = \frac{2\sigma^4}{n}$$

$$\mu_3(s^2) = \frac{8\sigma^6}{n^2} \qquad (148)$$

$$\mu_4(s^2) = \frac{12(n+4)\sigma^8}{n^3} \quad .$$

Clearly μ_3 and μ_4 are $0(\frac{1}{n})$ as $n \to \infty$ and $g(\sigma^2) = (\sigma^2)^2$. Therefore the theorem implies that, if a stabilizing transformation exists, it will be

$$f(s^2) = \int \frac{ds^2}{\sqrt{2}s^2} = \frac{\ell n\ s^2}{\sqrt{2}} + C = \sqrt{2}\ \ell n\ s + C \quad . \qquad (149)$$

For convenience, we may take C = 0 and consider the transformation $y = \ell n\ s^2 = 2.30259 \log_{10} s^2$. Since approximately

$$E(y') = \frac{1}{\sigma^2}$$

$$E(y'') = -\frac{1}{\sigma^4} \qquad (150)$$

$$[E(y'')]^2 = \frac{1}{\sigma^8} \quad ,$$

the type of reasoning employed in connection with the preceding example shows that

$$V(\ell n\ s^2) = \frac{2}{n} + 0(\frac{1}{n}) \quad \text{as } n \to \infty \text{ for fixed } \sigma^2 , \qquad (151)$$

$$V(\ell n\ s^2) = \frac{2}{n} + 0(\frac{1}{\sigma^2}) \quad \text{as } \sigma^2 \to \infty \text{ for fixed } n . \qquad (152)$$

Consequently $V(\ell n\ s^2) = \frac{2}{n}$ approximately, provided either n or σ^2 is sufficiently large.

(c) <u>Correlation coefficient computed for normal samples:</u>

Let $\theta = \rho$ (the product moment coefficient of correlation in a bivariate normal population) and let $T = r$ (a correlation coefficient computed from a random sample of n observations). The mean, variance, and higher moments of r are complicated functions of ρ and n but when n is large and $|\rho| \ll 1$,

$$
\begin{aligned}
E(r) &= \rho && \text{approximately.}\\
V(r) &= \frac{(1-\rho^2)^2}{n} && \text{approximately.}\\
\mu_3(r) &= 0(\tfrac{1}{n}) && \text{as } n \to \infty.\\
\mu_4(r) &= 0(\tfrac{1}{n}) && \text{as } n \to \infty.
\end{aligned}
\tag{153}
$$

Consequently the theorem (2.7.1) suggests the use of the following stabilizing transformation:

$$
\begin{aligned}
f(r) &= \int \frac{dr}{1-r^2}\\
&= \tfrac{1}{2}\ \ell n \frac{1+r}{1-r} + c \quad .
\end{aligned}
\tag{154}
$$

Taking c = 0, this yields

$$Z = \tfrac{1}{2}\ \ell n \frac{1+r}{1-r} = \tanh^{-1} r \quad . \tag{155}$$

The transformation of r was introduced by R. A. Fisher and will be discussed in a later section.

2.8 Anscombe's transformations of the Poisson, Binomial, and negative Binomial distributions:

2.8.1 Poisson variate: We consider now a general type of square root transformation $y = \sqrt{r+c}$ where r is a Poisson variate with mean m and c = const. Let t = r-m and m+c = m', then

$$y = (t+m+c)^{\frac{1}{2}} = (t+m')^{\frac{1}{2}}$$
$$= (m')^{\frac{1}{2}} \left(1 + \frac{t}{m'}\right)^{\frac{1}{2}} \quad .$$

Defining now the coefficient a_s as

$$a_s = (-1)^{s+1} \frac{1(-1)(-3)\ldots(-2s+3)}{2^s \, s!} \quad , \tag{156}$$

then for $t \geqq -m'$ we have the Taylor series expansion,

$$y = \sqrt{m'} \, \{1 + a_1 \frac{t}{m'} - a_2 \left(\frac{t}{m'}\right)^2 \ldots (-1)^s a_{s-1} \left(\frac{t}{m'}\right)^{s-1}\} + R_s \quad . \tag{157}$$

The remainder in Lagrange form is

$$R_s = \sqrt{m'} \, a_s \left(\frac{t}{m'}\right)^s \left(1+\theta \frac{t}{m'}\right)^{\frac{1}{2}-s} \quad , \quad 0 < \theta < 1 \quad .$$

This means that

$$|R_s| < \frac{a_s t^s}{(m')^{s-\frac{1}{2}}} \tag{158}$$

provided $t > 0$. Considering now $|t| \leqq m'$, we have directly from (157),

$$R_s(m')^{-\frac{1}{2}} = (1+\frac{t}{m'})^{\frac{1}{2}}-\{1+a_1\frac{t}{m'}\ldots+(-1)^s\ a_{s-1}(\frac{t}{m'})^{s-1}\}$$

$$= \sum_{i=s}^{\infty} (-1)^{i+1}\ a_i(\frac{t}{m'})^i \quad . \tag{159}$$

This series converges because $|\frac{t}{m'}| < 1$ and $(\frac{t}{m'})^i \to 0$ as $i \to \infty$. For the series a_i, let us take $\frac{a_{i+1}}{a_i} = \frac{\frac{1}{2}-i}{i+1} = \frac{\frac{1}{2}+1}{i+1} - 1$, $i > \frac{1}{2}$, so $|\frac{a_{i+1}}{a_i}| = 1 - \frac{3/2}{i+1}$ and $|\frac{a_i}{a_{i+1}}| = (1 - \frac{3/2}{i+1})^{-1}$ but $(1 - \frac{3/2}{i+1})^{-1} = 1 + (\frac{3/2}{i+1}) + (\frac{3/2}{i+1})^2(1+\varepsilon)$. Now

$\lim_{i\to\infty} i(|\frac{a_i}{a_{i+1}}|-1) = \lim_{i\to\infty} i\ \frac{3/2}{i+1} + \lim_{i\to\infty} i(\frac{3/2}{i+1})^2(1+\varepsilon)$. The second term $\to 0$, the first term $\to \frac{3}{2}$, and so $\lim_{i\to\infty} i(|\frac{a_i}{a_{i+1}}| - 1) > 1$ and the series $\Sigma|a_i|$ is convergent. Consequently $\Sigma\ a_i$ is convergent and thus the series $\sum_{i=s}^{\infty}(-1)^{i+1}\ a_i(\frac{t}{m'})^i$ converges. We may write

$$\frac{R_s(m')^{s-\frac{1}{2}}}{t^s} = \sum_{i=s}^{\infty} (-1)^{i+1}\ a_i(\frac{t}{m'})^{i-s} \quad , \tag{160}$$

where the right-hand side again converges and is therefore bounded. If G(s) is a bound to its absolute magnitude, then

$$|R_s| \leq G(s)\ \frac{|t|^s}{(m')^{s-\frac{1}{2}}} \quad . \tag{161}$$

Comparing this inequality with (158) we see that it holds for all $t > -m'$. We note now that the moments of t are:

$$\mu_1 = 0,\ \mu_2 = m,\ \mu_3 = m,\ \mu_4 = 3m^2+m,\ \text{etc.}, \tag{162}$$

and the absolute moment of order n is $O(m^{\frac{1}{2}n})$ as $m \to \infty$. We may therefore take expectations formally in the right-hand side of (157) and its powers, and derive asymptotic expansions for the moments of y as $m \to \infty$. We find:

$$E(y) = \sqrt{m'}\ \{1 - \frac{m}{8(m+c)^2} + \frac{m}{16(m+c)^3} - \frac{15m^2+5m}{128(m+c)^4} \ldots\}$$

$$= \sqrt{m+c} - \frac{m}{8(m+c)^{3/2}} + \frac{m}{16(m+c)^{5/2}} - \frac{15m^2+5m}{128(m+c)^{7/2}} + \ldots$$

$$= \sqrt{m+c} - \frac{1}{8m^{1/2}}(1+\frac{c}{m})^{-3/2} + \frac{1}{16m^{3/2}}(1+\frac{c}{m})^{-5/2} - \frac{15m+5}{128m^{5/2}}(1+\frac{c}{m})^{-7/2} + \ldots$$

$$= \sqrt{m+c} - \frac{1}{8m^{1/2}}\ (1 - \frac{3}{2}\frac{c}{m}) + \frac{1}{16m^{3/2}} - \frac{15}{128m^{3/2}} \ldots ,$$

ignoring terms of order more than $m^{-3/2}$. Hence

$$E(y) \sim \sqrt{m+c} - \frac{1}{8m^{1/2}} + \frac{3c+1}{16m^{3/2}} - \frac{15}{128m^{3/2}} = \sqrt{m+c} - \frac{1}{8m^{1/2}} - \frac{24c-7}{128m^{3/2}}\ . \tag{163}$$

If we set

$$E(y) = \sqrt{m_y+c}\quad , \tag{164}$$

then m_y is the estimate of m derived by applying the transformation $y = \sqrt{r+c}$ in reverse to the arithmetic mean $\bar{y}$ of a large sample of observed values of y. From (163) we have

$$m_y = [E(y)]^2 - c$$

$$= m+c+\frac{1}{64m} - \frac{(m+c)^{\frac{1}{2}}}{4m^{1/2}} + \frac{(24c-7)(m+c)^{\frac{1}{2}}}{64m^{3/2}}$$

$$= m + \frac{1}{64m} - \frac{1}{4}(1 + \frac{c}{m})^{\frac{1}{2}} + \frac{24c-7}{64m}(1 + \frac{c}{m})^{\frac{1}{2}}$$

$$= m + \frac{1}{64m} - \frac{1}{4} - \frac{1}{8}\frac{c}{m} + \frac{24c-7}{64m} = m - \frac{1}{4} + \frac{8c-3}{32m} \quad , \qquad (165)$$

so that setting $c = \frac{3}{8}$ renders the bias $m_y = m$ in m_y nearly constant.

For the higher moments we have

$$E(y^2) = E(r+c) = m+c \quad , \qquad (166)$$

$$E(y^3)=E(y^2y)=E[(r+c)y]=E[(r-m+m+c)y]=e[(t+m')y]=m'E(y)+E(ty) \quad .$$

$$E(y^4)=E(r+c)^2=V(r+c)+[E(r+c)]^2=m+(m+c)^2 \,,$$

after computing μ_1', μ_2', μ_3', μ_4' for the variable y. We deduce that

$$V(y) = \mu_2' - \hat{\mu}^2 \sim \frac{1}{4}\{1+\frac{3-8c}{8m}+\frac{32c^2-52c+17}{32m^2}\} \quad . \qquad (167)$$

In particular for $c = \frac{3}{8}$ we have

$$Var(y) \sim \frac{1}{4}\{1 + \frac{1}{16m^2}\} \quad . \qquad (168)$$

For the skewness and kurtosis of the distribution of y we have

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \sim -\frac{1}{2m^{1/2}}\{1 + \frac{25-48c}{16m}\} \quad , \qquad (169)$$

compared with $m^{-\frac{1}{2}}$ for the original Poisson r. Similarly

$$\gamma_2 = \frac{\mu_4}{\mu_2^{\,2}} - 3 \sim \frac{1}{m}\{1 + \frac{945-1536c}{256m}\} \quad , \tag{170}$$

and this compares with $\gamma_2 = m^{-1}$ for the original r. It is to be noticed that although the transformation with $c = \frac{3}{8}$ renders more stabilization for the variance, it gives no improvement over the original Poisson variable with regard to normality.

It is also of interest to find the large sample efficiency E_y of the arithmetic mean $\bar{y}$ of observed values of y as a statistic for determining m. If X is a random variable having a distribution such that the arithmetic mean $\bar{X}$ is a sufficient statistic for determining a parameter θ, the large sample efficiency E_Y of the average $\bar{Y}$ of any function Y of X, for determining θ, is the square of the correlation coefficient between X and Y, i.e.,

$$E_Y = \frac{[\mathrm{Cov}(X,Y)]^2}{\mathrm{Var}(X)\cdot \mathrm{Var}(Y)} \quad . \tag{171}$$

In the present case we have

$$\mathrm{Cov}(r,y)=E(ry)-E(r)E(y)=E[ty+my]-mE(y)=E(ty) \quad ,$$

and

$$E_y \sim 1 - \frac{1}{8m} + \frac{16c-9}{64m^2} \quad .$$

For $c = \frac{3}{8}$ we have,

$$E_y \sim 1 - \frac{1}{8m} - \frac{3}{64m^2} \quad . \qquad (172)$$

The foregoing results were published by Anscombe (1948).

2.8.2 <u>Binomial variate</u>: We suppose now that r is distributed as a Binomial distribution with index n and mean (0 < m < n). We consider the transformation

$$y = \sqrt{n+d_2} \quad \sin^{-1}\sqrt{\frac{r+c}{n+d_1}} \quad , \qquad (173)$$

where c, d_1, and d_2 are constants to be determined. Setting $r-m = t$, $m+c = m'$, $n+d_1 = n_1$, $n+d_2 = n_2$, the transformation becomes

$$y = \sqrt{n_2} \quad \sin^{-1}\sqrt{\frac{m'+t}{n_1}} \quad . \qquad (174)$$

The expansion of $\frac{dy}{dt}$ gives a Binomial series which is convergent in the range $|\frac{m'+t}{n_1}| < 1$, and thus the expansion of y in (174) in ascending powers of t is also convergent in the same range. Since we need $m'+t > 0$, the series for y is convergent in the range $0 < \frac{m'+t}{n_1} < 1$ or $-m' < t < n_1-m'$.
The series is,

$$y = \sqrt{n_2} \; \{(\frac{m'+t}{n_1})^{1/2} + \frac{1}{6}(\frac{m'+t}{n_1})^{3/2} + \frac{3}{40}(\frac{m'+t}{n_1})^{5/2} + \ldots\}$$

In analogy with the Poisson case, we can obtain asymptotic expansions for the moments of y, valid for large n and constant ratio $\frac{m}{n}$. Anscombe obtained the variance of y as

$$\mathrm{Var}(y) = \frac{1}{4}\{1 + \frac{2d_2-1}{2n} + \frac{3-8c}{8m} + \frac{3+8c-8d_1}{8(n-m)}\} \quad . \tag{175}$$

So that if we choose $c = \frac{3}{8}$, $d_1 = \frac{3}{4}$, $d_2 = \frac{1}{2}$, we have

$$\mathrm{Var}(y) = \frac{1}{4} + 0(\frac{1}{n_2}) \quad , \tag{176}$$

where $$y = \sqrt{n+\tfrac{1}{2}}\ \sin^{-1}\sqrt{\frac{r+3/8}{n+3/4}} \quad . \tag{177}$$

It will be noticed here that the choice of d_2 only effects the scale of y (for n fixed), and not the constancy of Var(y) as n varies, nor the shape of the distribution of y. The quantities γ_1, γ_2, E_y are

$$\gamma_1 \sim \frac{2m-n}{2[nm(n-m)]^{\frac{1}{2}}} = \frac{P-Q}{2(nPQ)^{\frac{1}{2}}} \tag{178}$$

$$\gamma_2 \sim \frac{n^2-2m(n-m)}{nm(n-m)} = \frac{1-2PQ}{nPQ} \tag{179}$$

$$E_y \sim 1 - \frac{(2m-n)^2}{8nm(n-m)} = 1 - \frac{(P-Q)^2}{8nPQ} \quad , \tag{180}$$

where $P = \frac{m}{n}$, $Q = 1-P$.

It should be noted that the nearer P is to a half, the better the transformation becomes.

2.8.3 Negative Binomial variate: Consider a Negative Binomial variate r with mean m and exponent K $(m,k > 0)$ such that the probability of observing a value r is

$$p_r = \frac{\Gamma(r+k)}{r!\Gamma(k)} \left(\frac{m}{m+k}\right)^r \left(1 + \frac{m}{k}\right)^{-k} \quad (r=0,1,2 \ldots). \qquad (181)$$

Anscombe gives, in analogy with the Binomial case, the transformation

$$y = \sqrt{k-\tfrac{1}{2}} \sinh^{-1}\sqrt{\frac{r+3/8}{k-3/4}} \qquad (182)$$

valid for large m and constant ratio $\frac{k}{m}$, with

$$\text{Var}(y) = \frac{1}{4} + 0\left(\frac{1}{m^2}\right) \quad . \qquad (183)$$

However, it is of more interest to consider m large but k fixed. (The corresponding problem does not arise with the positive Binomial, since $m < n$ necessarily.) Anscombe consideres two transformations,

$$y = 2 \sinh^{-1}\sqrt{\frac{r+c}{k+d}} \qquad (184)$$

$$y = \ell n(r+A) \quad . \qquad (185)$$

It is supposed that c, k+d, and A are positive and constant. Apart from an added constant (184) may be written

$$y = 2\, \ell n[\sqrt{r+c} + \sqrt{(r+c+K+d)}] \quad . \qquad (186)$$

When r is large, we have (again ignoring an added constant):

$$y = 2\ \ell n[\sqrt{r+c} + \sqrt{r+c+K+d}\,] = 2\ \ell n\{\sqrt{r}\,[(1+\frac{c}{r})^{\frac{1}{2}}+(1+\frac{c+k+d}{r})^{\frac{1}{2}}]\}$$

$$\sim 2\ \ell n\{\sqrt{r}\,[1+\frac{c}{2r}-\frac{1}{8}\,\frac{c^2}{r^2}+1+\frac{c+k+d}{2r}-\frac{1}{8}\,\frac{(c+k+d)^2}{r^2}]\}$$

$$= 2\ \ell n\sqrt{r} + 2\ \ell n[2 + \frac{2c+k+d}{2r} - \frac{2c^2+2c(k+d)+(k+d)^2}{8r^2}]$$

$$= \ell n\ r + 2\ \ell n[1 + \frac{2c+k+d}{4r} - \frac{2c^2+2c(k+d)+(k+d)^2}{16r^2}] + \ell n\ 4\ ,$$

ignoring $\ell n\ 4$ we get,

$$y \sim \ell n\ r + \frac{2c+k+d}{2r} - \frac{2c^2+2c(k+d)+(k+d)^2}{8r^2} - \frac{(2c+k+d)^2}{16r^2}$$

$$= \ell n\ r + \frac{A}{r} - \frac{B^2}{2r^2}\ , \qquad (187)$$

where

$$A = \frac{1}{2}(2c+k+d)\ ,$$

$$B^2 = \frac{1}{8}[8c^2+8c(k+d)+3(k+d)^2]\ . \qquad (188)$$

For (185) $\qquad B = A\ .$

We proceed to find an asymptotic expansion (as $n \to \infty$, with K fixed) for the moment generating function of y, i.e., for:

$$M(t) \equiv \sum_{r=0}^{\infty} e^{yt}\, P_r \qquad (189)$$

we set,

$$\frac{m}{m+k} = e^{-\alpha}\ , \qquad (190)$$

so that $\alpha \to 0$ as $m \to \infty$; and

$$e^{yt} \frac{\Gamma(r+k)}{\Gamma(k)r!} e^{-r\alpha} = u_r(\alpha) \quad . \tag{191}$$

We state now the following lemma without proof:

<u>Lemma</u>: As $\alpha \to 0$

$$\sum_{r=0}^{\infty} u_r(\alpha) - \int_0^{\infty} u_r(\alpha)\,dr \tag{192}$$

tends to a finite limit (depending on k and t, and on which function y or r is chosen, namely (184) or (185).

Applying this lemma to (189) we have:

$$M(t) = \sum_{r=0}^{\infty} e^{yt} \frac{\Gamma(k+r)}{\Gamma(k)r!} e^{-\alpha r}(1-e^{-\alpha})^k = (1-e^{-\alpha})^k \sum_{r=0}^{\infty} u_r(\alpha)$$

$$= (1-e^{-\alpha})^k \int_0^{\infty} u_r(\alpha)\ dr + 0(\alpha^k) \quad . \tag{193}$$

In general we cannot evaluate this integral exactly. We expand $u_r(\alpha)$ for large r and find the first term to be $\frac{r^{t+k-1}\ e^{-r\alpha}}{\Gamma(k)}$. The error of the expansion is always less than a multiple of the next term (independent of α) for $r \geqq 1$. Integrating term by term between the limits 1 and ∞, we obtain the following:

<u>Theorem</u>: With the definitions of (181), (187), (189), and (190), M(t) can be expanded asymptotically for $\alpha \to 0$

in the form:

$$M(t) = \frac{\Gamma(k+t)}{\alpha^t \Gamma(k)}[1+(A-\tfrac{1}{2}k)t\frac{\alpha}{k+t-1}+\{(\tfrac{1}{2}(A-\tfrac{1}{2}k)^2+\frac{1}{24}k)t^2 \tag{194}$$

$$+(\tfrac{1}{2}kA-\frac{1}{24}k(k+3)-\tfrac{1}{2}B^2)t\}\frac{\alpha^2}{(k+t-1)(k+t-2)}+\ldots]+0(\alpha^k) .$$

The series in square brackets is continued as far as the term in α^n, where n is the greatest integer less than k, t is supposed confined to a neighbourhood of zero.

The cumulant generating function is

$$K(t) = \ell n\ M(t) = -t\ \ell n\ \alpha + \ell n\ \Gamma(k+t) - \ell n\ \Gamma(k)$$

$$+ (A-\tfrac{1}{2}k)t\ \frac{\alpha}{k+t-1} + \ldots \quad , \tag{195}$$

(by taking ℓn M(t) and expanding ℓn of the square brackets in (194) as $\ell n(1+x) = x - \frac{x^2}{2} \ldots$).

For Var(y) we need K_2. Now

$$\ell n\ \Gamma(t+k) = \ell n\ \Gamma(k) + \frac{t}{1!}\ \Psi(k) + \frac{t^2}{2!}\ \Psi'(k) \ \ldots \quad ,$$

where $\Psi(k)$, $\Psi'(k)$, etc. denote the successive derivatives of $\ell n\ \Gamma(k)$. Also we have

$$(A-\tfrac{1}{2}k)\alpha t(k+t-1)^{-1} = \frac{2A-k}{2}\ \alpha t(k-1)^{-1}(1+\frac{t}{k-1})^{-1}$$

$$= \frac{2A-k}{2(k-1)}\ \alpha t(1-\frac{t}{k-1}+\ldots)$$

$$= \frac{k-2A}{(k-1)^2}\ \alpha\ \frac{t^2}{2} + \ldots \quad .$$

Hence

$$K_2 = \mathrm{Var}(y) \sim \Psi'(k) + \frac{k-2A}{(k-1)^2}\alpha \qquad (k > 1) \quad . \qquad (196)$$

If $A = \frac{1}{2}k$, we have

$$K(t) = -t\ \ell n\ \alpha + \ell n\Gamma(k+t) - \ell n\Gamma(k) + \{\frac{t^2k}{24} + (\tfrac{1}{4}k^2 - \frac{1}{24}k(k+3) - \tfrac{1}{2}B^2)t\}$$

$$\cdot \frac{\alpha^2}{(k-1)(k-2)(1+\frac{t}{k-1})(1+\frac{t}{k-2})} \quad ,$$

in which the last term equals

$$\{\frac{t^2k}{24(k-1)(k-2)} - \frac{5k^2-3k-12B^2}{24k(k-1)(k-2)}t\}\alpha^2(1 - \frac{t(2k-3)}{(k-1)(k-2)} + \ldots) \quad .$$

For the coefficient of $\frac{t^2}{2}$ we get

$$\frac{k(k-1)(k-2)-(2k-3)(5k^2-3k-12B^2)}{12(k-1)^2(k-2)^2}\alpha^2 \quad ,$$

and so

$$\mathrm{Var}(y) \sim \Psi'(k) + \frac{k(k-1)(k-2)-(2k-3)(5k^2-3k-12B^2)}{12(k-1)^2(k-2)^2}\alpha^2 \ ,$$

$$(k>2) \ . \quad (197)$$

Considering y defined by (184), the condition $A = \frac{1}{2}k$ gives $d = -2c$, and the coefficient of α^2 in Var(y) vanishes if c takes a value dependent on k, which for large k is approximately

$$c = \frac{3}{8} + \frac{23}{192}\frac{1}{k} \quad , \qquad (198)$$

and which rises to a little above 0.4 as k decreases towards 2. For y defined by (185), we set $A = \frac{1}{2}k$. In this case $B = A = \frac{1}{2}k$, so substituting B in (197) and simplifying the results we get

$$\mathrm{Var}(y) \sim \Psi'(k) - \frac{k(3k^2-9k+7)}{12(k-1)^2(k-2)^2}\,\alpha^2 \quad (k > 2) \quad . \tag{199}$$

If k is large and $m \gg k$, we have $\alpha \sim \frac{k}{m}$ and

$$\mathrm{Var}(y) \sim \frac{1}{k}\left(1 - \frac{k^2}{4m^2}\right) \quad . \tag{200}$$

Thus the larger k is the larger m must be for Var(y) to approach its limiting value when $m \to \infty$. The transformation (185) is therefore not satisfactory if k is large.

For either form of transformation, Anscombe gives the following limiting values as $m \to \infty$ $(\alpha \to 0)$:

$$\begin{aligned} m_y/m &= \exp\{\Psi(k) - \ell n\ k\} \\ \gamma_1 &= \Psi''(k)/[\Psi'(k)]^{3/2} \\ \gamma_2 &= \Psi'''(k)/[\Psi'(k)]^2 \quad . \end{aligned} \tag{201}$$

3. Transformations for solving the tail and confidence interval problems in Binomial, Poisson, χ^2, and negative Binomial.

3.1 Introduction: In this section, the main idea is to derive a transformation of the distribution under consideration which can be used to facilitate the computation of tail sums of the distribution, that is the calculation of the probability that the variable is greater than or equal to a given value a, or less than or equal to a, and to provide also a numerical determination of confidence limits for an unknown parameter in terms of an observed value a. We can formulate these two problems in compact form as follows:

Let x be a random variable. By $p(x \geq a; \theta)$ we denote the probability that the variable is greater than or equal to a, where θ is a parameter (known or unknown). $p(x \geq a; \theta)$ is assumed to be a monotonic function of θ for any fixed a. If we put

$$p(x \geq a; \theta) = \varepsilon \quad , \tag{202}$$

the tail problem arises when ε is to be calculated, and the confidence problem if θ is unknown. In the latter case $1-\varepsilon$ is the confidence level of a one-sided interval.

3.2 Binomial case: Let x follow the Binomial distribution with the parameters n and p. Further, let the variable y be Beta-distributed with f_1 and f_2 degrees of freedom so that y has the frequency function:

$$\text{const. } y^{\frac{1}{2}f_1-1}(1-y)^{\frac{1}{2}f_2-1} \quad .$$

The two distributions are connected by means of the well-known formula

$$p^n+np^{n-1}q+\ldots+\binom{n}{a}p^a q^{n-a}=\frac{\Gamma(n+1)}{\Gamma(a)\Gamma(n-a+1)}\int_0^p y^{a-1}(1-y)^{n-a}dy \tag{203}$$

or

$$p(x \geqq a;\ p) = p(y \leqq p;\ f_1,\ f_2) \quad , \tag{204}$$

where

$$f_1 = 2a, \quad f_2 = 2(n-a+1) \quad . \tag{205}$$

Using the fact that the variable $\frac{f_1 F}{f_1 F+f_2}$ has the β-distribution with f_1, f_2 degrees of freedom, we can write

$$P(x \geqq a;\ p) = P\left(\frac{f_1 F}{f_1 F+f_2} \leqq p,\ f_1,\ f_2\right) \quad ,$$

where f_1, f_2 are given in (205) with

$$p^* = \frac{a}{n+1}\ , \quad q^* = 1-p^* = \frac{n-a+1}{n+1} \quad . \tag{206}$$

We find

$$P(x \geqq a; p) = P\left(\frac{f_1 F}{f_2} \leqq \frac{p}{1-p}; f_1, f_2\right) = P\left(\frac{p^*}{q^*} F \leqq \frac{p}{q}\right) = P(z \leqq Z) \quad , (207)$$

where $\frac{f_1}{f_2} = \frac{p^*}{q^*}$ and z has the Fisher z-distribution with the degrees of freedom given by (205).

For the value of Z, we can write (207) as

$$P\left(F \leqq \frac{p}{q} / \frac{p^*}{q^*}\right) = P\left(\tfrac{1}{2} \log F \leqq \tfrac{1}{2} \log \frac{p}{q} - \tfrac{1}{2} \log \frac{p^*}{q^*}\right) = P(z \leqq Z) \quad .$$

Hence

$$2Z = \log \frac{p}{q} - \log \frac{p^*}{q^*} \quad . \qquad (208)$$

It follows from these relations that the confidence problem for a Binomial variable is identical with a percentage point problem for a z-variable. In addition, we have replaced a discontinuous variable by a continuous one, which has certain advantages.

3.2.1 <u>The Cornish-Fisher expansion for the z-distribution</u>:

We shall now derive an important expansion by using the Cornish-Fisher expansion of the z-distribution. This series runs as follows:

$$P(z \leqq Z; f_1, f_2) = \phi(\lambda) \quad , \qquad (209)$$

where

$$Z = \lambda\sqrt{\frac{\sigma}{2}} - \frac{\delta}{6}(\lambda^2+2)+\sqrt{\frac{\sigma}{2}}\left[\frac{\sigma}{24}(\lambda^3+3\lambda)+\frac{\delta^2}{72\sigma}(\lambda^3+11\lambda)\right]+\ldots \qquad (210)$$

and $\phi(\lambda) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\lambda} e^{-\frac{1}{2}u^2}du$ is the normal distribution function.

Here for brevity, $\sigma = \frac{1}{f_1} + \frac{1}{f_2}$ and $\delta = \frac{1}{f_1} - \frac{1}{f_2}$. Inserting the values (205) of f_1, f_2 and using the notation (206), we find:

$$\sigma = \frac{1}{2(n+1)}\,\frac{1}{p^*q^*}\,, \qquad \delta = \frac{1}{2(n+1)}\,\frac{q^*-p^*}{p^*q^*} \quad .$$

We insert these expressions in (210) and substitute the right member of (208) for 2Z. Combining (207) and (209) and changing the sign of λ, in order to get the upper tail sum of the Binomial in terms of the upper tail of the normal, we have finally

$$p(x \geq a;\ p) = 1 - \phi(\lambda) \quad , \qquad (211)$$

where a, p, and λ satisfy the relation:

$$2Z = \log\left(\frac{p}{q}\right) - \log\frac{p^*}{q^*} = c_1(n+1)^{-\frac{1}{2}} + c_2(n+1)^{-1} + c_3(n+1)^{-\frac{3}{2}} \quad . \qquad (212)$$

We write:

$$2Z = -2\lambda\sqrt{\frac{\sigma}{2}-\frac{\delta}{3}(\lambda^2+2)}+\sqrt{\frac{\sigma}{2}}\,[\frac{\sigma}{12}(-\lambda^3-3\lambda)+\frac{\delta^2}{36\sigma}(-\lambda^3-11\lambda)]+\ldots$$

$$= -2\lambda\sqrt{\frac{1}{4p^*q^*(n+1)} - \frac{1}{6}\,\frac{q^*-p^*}{p^*q^*(n+1)}(\lambda^2+2)}+\sqrt{\frac{1}{4p^*q^*(n+1)}}$$

$$\cdot\ [\frac{1}{24(n+1)p^*q^*}(-\lambda^3-3\lambda) + \frac{(q^*-p^*)^2}{144p^{*2}q^{*2}(n+1)^2}2p^*q^*(n+1)$$

$$\cdot\ (-\lambda^3-11\lambda)] + \ldots$$

$$= \frac{-\lambda}{\sqrt{p^*q^*}}(n+1)^{-\frac{1}{2}} - \frac{(q^*-p^*)(\lambda^2+3)}{6p^*q^*}(n+1)^{-1}+\frac{1}{2\sqrt{p^*q^*}}\,(n+1)^{-\frac{1}{2}}$$

$$\cdot\ [\frac{-(\lambda^3+3\lambda)}{24p^*q^*}(n+1)^{-1} - \frac{(q^*-p^*)^2(\lambda^3+11\lambda)}{72p^*q^*}(n+1)^{-1}] + \ldots$$

$$\equiv c_1(n+1)^{-\frac{1}{2}}+c_2(n+1)^{-1}+c_3(n+1)^{-\frac{3}{2}} \quad , \quad \text{where}$$

$$c_1 = -\lambda(p^*q^*)^{-\frac{1}{2}}; \quad c_2 = -\frac{q^*-p^*}{6p^*q^*}(\lambda^2+2) = \frac{\lambda^2+2}{6}(p^*-q^*)(p^*q^*)^{-1} ;$$

$$c_3 = \frac{-(\lambda^3+3\lambda)}{48}(p^*q^*)^{-\frac{3}{2}} - \frac{\lambda^3+11\lambda}{144}(p^*q^*)^{-\frac{3}{2}}(q^*-p^*)^2$$

$$= \frac{-(\lambda^3+3\lambda)}{48}(p^*q^*)^{-\frac{3}{2}} - \frac{\lambda^3+11\lambda}{144}(p^*q^*)^{-\frac{3}{2}}(1-4p^*q^*)$$

$$= (-\frac{\lambda^3+3\lambda}{48} - \frac{\lambda^3+11\lambda}{144})(p^*q^*)^{-\frac{3}{2}} + \frac{\lambda^3+11\lambda}{36}(p^*q^*)^{-\frac{1}{2}}$$

$$= (\frac{-4\lambda^3-20\lambda}{144})(p^*q^*)^{-\frac{3}{2}} + \frac{\lambda^3+11\lambda}{36}(p^*q^*)^{-\frac{1}{2}}$$

$$= -\frac{\lambda^3+5\lambda}{36}(p^*q^*)^{-\frac{3}{2}} + \frac{\lambda^3+11\lambda}{36}(p^*q^*)^{-\frac{1}{2}} \quad .$$

3.2.2 <u>The Cornish-Fisher expansion for a transformed</u>

<u>Binomial variable</u>: We shall apply the Cornish-Fisher method, not to $\log \frac{p}{q} - \log \frac{p^*}{q^*}$ but to a general function $\phi(p) - \phi(p^*)$. The function $\phi(x)$ is assumed to define a one to one correspondence between points in some interval $0 < a \leq x \leq b < 1$ on the x-axis and a corresponding interval on the ϕ-axis. Further, $\phi(x)$ is supposed to have bounded derivatives of the inverse function $p = p(\phi)$ with respect to ϕ (which are also supposed to be bounded). The value $\phi(p^*)$ will sometimes for brevity be denoted by ϕ^*.

We shall replace (212) by two other expansions. The first expansion can be written in the form:

$$\phi(p) = \phi(p^*) + a_1(n+1)^{-\frac{1}{2}} + a_2(n+1)^{-1} + a_3(n+1)^{-\frac{3}{2}} + \ldots \quad (213)$$

in which the coefficients a_ν depend upon λ and p^*, not upon p. We shall call this series the first transformation series.

The second expansion is of the same form

$$\phi(p^*) = \phi(p) + b_1(n+1)^{-\frac{1}{2}} + b_2(n+1)^{-1} + b_3(n+1)^{-\frac{3}{2}} + \ldots, \quad (214)$$

where the coefficients b_ν depends only upon λ and p, not upon p*. We call this the second transformation series.

We begin by determining the coefficients a_ν. Expanding the left member of (212) formally in Taylor series around ϕ^*, we find:

$$\log\frac{p}{q}-\log\frac{p^*}{q^*}=f(\phi)-f(\phi^*)=d_1^*(\phi-\phi^*)+\frac{d_2^*}{2!}(\phi-\phi^*)^2+\ldots, \quad (215)$$

where d_i^* is d_i after replacing p by p*. Now $d_1 = f'_\phi(\phi)$, $d_2 = f''_\phi(\phi)$, etc. Thus

$$d_1 = \frac{d}{d\phi}(\log\frac{p}{q}) = \frac{p'}{pq} = \frac{1}{\phi'_p pq} \quad (216)$$

because $p'_\phi \cdot \phi'_p = 1$, and

$$d_2 = \frac{d^2}{d\phi^2}(\log\frac{p}{q}) = \frac{d}{d\phi}(\frac{p'}{pq}) = \frac{p''}{pq} + \frac{p'^2(p-q)}{p^2q^2} \quad .$$

But the relation $p'_\phi \cdot \phi'_p = 1$ gives

$$p''_\phi \cdot \phi'_p + p'_\phi(\phi'_p)'_\phi = 0$$

or

$$p''_\phi \cdot \phi'_p + p'_\phi(\phi'_p)'_p \cdot p'_\phi = p''_\phi \cdot \phi'_p + p'^2_\phi \cdot \phi''_p = 0$$

or

$$p''_\phi = (-p'^2_\phi \cdot \phi''_p)/\phi'_p = -\phi''/\phi'^3 \quad ,$$

where $(\phi'_p)'_\phi$ means the derivative of ϕ'_p with respect to ϕ. Hence

$$d_2 = -\frac{1}{\phi'^2 pq}(\frac{\phi''}{\phi'} + \frac{q-p}{pq}) \quad , \quad (217)$$

and similarly for d_3, d_4, etc.

We determine a_1, a_2 successively by inserting the series (213) in (215) and comparing the coefficients of $(n+1)^{-\frac{1}{2}\nu}$ in the right member of (212) and (215). We find

$$\log\frac{p}{q} - \log\frac{p^*}{q^*} = c_1(n+1)^{-\frac{1}{2}}+c_2(n+1)^{-1}+\dots$$

$$\doteq d_1^*a_1(n+1)^{-\frac{1}{2}}+d_1^*a_2(n+1)^{-1}+\frac{d_2^*}{2!}a_1^2(n+1)^{-1}+\dots \quad .$$

Hence

$$a_1d_1^* = c_1\ , \qquad a_2d_1^* + \frac{a_1^2d_2^*}{2!} = c_2$$

or

$$a_1 = -\lambda(p^*q^*)^{\frac{1}{2}}\ \phi'(p^*)$$
$$a_2 = \tfrac{1}{2}\lambda^2[p^*q^*\phi''(p^*)+(1-K(\lambda))(q^*-p^*)\phi'(p^*)] \quad , \tag{218}$$

where

$$K(\lambda) = \frac{\lambda^2+2}{3\lambda^2} \quad . \tag{219}$$

The coefficients b_ν may be determined directly by a modification of the above procedure. We expand $\log\frac{p}{q}-\log\frac{p^*}{q^*}$ around ϕ instead of ϕ^*, yielding

$$2Z = \log\frac{p}{q}-\log\frac{p^*}{q^*} \doteq -d_1(\phi^*-\phi)-\frac{d_2}{2!}(\phi^*-\phi)^2+\dots \quad , \tag{220}$$

where d_1 and d_2 are given by (216) and (217). Further, the coefficients c_ν in (212) are also expanded around ϕ. After substituting (214) in the resulting series and in (220), the coefficients of $(n+1)^{-\frac{1}{2}\nu}$ are compared and we obtain

$$b_1 = \lambda(pq)^{\frac{1}{2}} \phi'(p) \quad ,$$
$$b_2 = \tfrac{1}{2}\lambda^2[pq\phi''(p)+K(\lambda)(q-p)\phi'(p)] \quad . \tag{221}$$

3.2.3 Determination of the 'best' transformation of a Binomial variable: In many practical situations only the first two terms in (213), (214) are calculated, and the terms of higher order are neglected. It is natural then to choose the transformation $\phi(p)$ in such a way that the term of order $(n+1)^{-1}$ vanishes or at least becomes very small. An application of this principle gives, of course, no guarantee that the total error will be small, but good results may be expected. The coefficient a_2 is given in (217) and may be written

$$a_2 = \tfrac{1}{2}\lambda^2 p^* q^*[\phi''(p^*) \ \alpha_1(\lambda)(\frac{1}{p^*} - \frac{1}{q^*})\phi'(p^*)] \quad ,$$

where

$$\alpha_1(\lambda) = 1 - K(\lambda) = 1 - \frac{\lambda^2+2}{3\lambda^2} \quad . \tag{222}$$

Putting $a_2 = 0$ and replacing p^* by p, we obtain the differential equation:

$$\frac{\phi''(p)}{\phi'(p)} = \alpha_1(\lambda)(\frac{1}{1-p} - \frac{1}{p}) \quad . \tag{223}$$

In terms of a general parameter α, this euqation has, apart from a multiplicative constant, the solution:

$$\frac{\phi''}{\phi'} = \alpha\left(\frac{1}{1-p} - \frac{1}{p}\right) \quad \text{or} \quad \log \phi' = -\alpha \log[p(1-p)] \quad .$$

Thus $[\phi' \; p^{\alpha}(1-p)^{\alpha}] = 1$

or
$$\phi = \int_{p_o}^{p} x^{-\alpha}(1-x)^{-\alpha} \, dx \quad . \tag{224}$$

Starting with b_2 in (221) and proceeding analogously, we find the same transformation with the parameter

$$\alpha_2(\lambda) = K(\lambda) = \frac{\lambda^2+2}{3\lambda^2} \quad . \tag{225}$$

We observe that $\alpha_1(\lambda) + \alpha_2(\lambda) = 1$. When λ varies from 0 to ∞, $\alpha_1(\lambda)$ increases from $-\infty$ to $\frac{2}{3}$ and $\alpha_2(\lambda)$ decreases from $+\infty$ to $\frac{1}{3}$. For $\lambda = 2$ we have $\alpha_1 = \alpha_2 = \frac{1}{2}$.

Because of the similarity of (224) with the Beta-integral, we shall call the transformation (224) the Beta-transformation. It is interesting to note that most of the transformations we have seen previously are special cases of this transformation. This will become clear later where special values will be assigned to α.

The expansions (213) and (214) can be simplified when the Beta-transformation is used. Substituting $\phi'(p^*)$, $\phi''(p^*)$ for the expressions obtained by differentiating (224), we have in general that if $\phi(\alpha) = \int_A^B f(x,\alpha)dx$ then

$$\frac{d\phi(\alpha)}{dB} = \int \frac{\partial}{\partial B}[f(x,\alpha)]dx + f(B,\alpha)\frac{dB}{dB} - f(A,\alpha)\frac{dA}{dB} .$$ Hence

$$\phi'(p^*) = [p^*(1-p^*)]^{-\alpha} = (p^*q^*)^{-\alpha}, \ \phi''(p^*) = -\alpha(p^*q^*)^{-\alpha-1}(q-p) .$$

Now

$$a_1 = -\lambda(p^*q^*)^{\frac{1}{2}}(p^*q^*)^{-\alpha} = -\lambda(p^*q^*)^{\frac{1}{2}-\alpha} ,$$

$$\begin{aligned} a_2 &= \tfrac{1}{2}\lambda^2\{p^*q^*[-\alpha(p^*q^*)^{-\alpha-1}q + \alpha(p^*q^*)^{-\alpha-1}p]\} + \frac{\lambda^2-1}{3}(q^*-p^*)(p^*q^*)^{-\alpha} \\ &= \tfrac{1}{2}\lambda^2\{p^*q^*[2\alpha(p^*q^*)^{-\alpha-1}p^* - \alpha(p^*q^*)^{-\alpha-1}]\} + \frac{\lambda^2-1}{3}(q^*-p^*)(p^*q^*)^{-\alpha} \\ &= \tfrac{1}{2}\lambda^2(p^*q^*)^{-\alpha}\alpha[p-q] + \frac{\lambda^2-1}{3}(q^*-p^*)(p^*q^*)^{-\alpha} \\ &= (p^*q^*)^{-\alpha}(q^*-p^*)[-\tfrac{1}{2}\lambda^2\alpha + \frac{\lambda^2-1}{3}] \\ &= \frac{(2-3\alpha)\lambda^2-2}{6}(q^*-p^*)(p^*q^*)^{-\alpha} . \end{aligned}$$

Hence after substituting in (213) we have

$$\phi(p,\alpha) = \phi(p^*,\alpha) - \lambda(p^*q^*)^{\frac{1}{2}-\alpha}(n+1)^{-\frac{1}{2}} + \frac{(2-3\alpha)\lambda^2-2}{6}(q^*-p^*)(p^*q^*)^{-\alpha} \cdot (n+1)^{-1} + \ldots \quad . \qquad (226)$$

If we solve with respect to λ in the second term of the right member, we obtain

$$\lambda = (\phi^*-\phi)(p^*q^*)^{\alpha-\frac{1}{2}}(n+1)^{\frac{1}{2}} + \frac{(2-3\alpha)\lambda^2-2}{6}(q^*-p^*)(p^*q^*)^{-\frac{1}{2}}(n+1)^{-\frac{1}{2}} + \ldots \quad . \qquad (227)$$

Hence we have $P(x \geq a, p) = 1 - \phi(\lambda)$ where λ is given in (227).

The second transformation series leads to

$$P(x \geqq a, p) = 1 - \phi(\lambda) \quad ,$$

where λ is given as

$$\lambda=(\phi^*-\phi)(pq)^{\alpha-\frac{1}{2}}(n+1)^{\frac{1}{2}}+\frac{(3\alpha-1)\lambda^2-2}{6}(q-p)(pq)^{-\frac{1}{2}}(n+1)^{-\frac{1}{2}}+\ldots \quad (228)$$

The quantities p^* and q^* which appear in the formulae have been defined in (206). It is seen that the part of the expansion (228) reproduced here is obtained from (227) by replacing p^*, q^* by p and q, and α in the second term of the right-hand member by $1-\alpha$.

If we want the third terms in the series (227) and (228), the coefficient of $(n+1)^{-\frac{3}{2}}$ is

$$\frac{\lambda}{36}(A - \frac{B}{pq}) \quad , \qquad (229)$$

where the values of A and B in the series (227) are

$$\begin{aligned} A &= [1+12(\alpha-1)(2\alpha-1)]\lambda^2+48\alpha-37 \\ B &= [1+6\alpha(\alpha-1)]\lambda^2+12\alpha-7 \quad , \end{aligned} \qquad (230)$$

and in the series (228), A and B become

$$\begin{aligned} A &= [1+12\alpha(2\alpha-1)]\lambda^2-48\alpha+11 \\ B &= (6\alpha^2-\tfrac{1}{2})\lambda^2-12\alpha-1 \quad . \end{aligned} \qquad (231)$$

We have studied so far the right (upper) tail of the Binomial distribution. The other tail is, of course, obtained by substituting $a+1$ for a and taking the complementary probability.

Thus

$$P(x \leqq a,\ p) = 1-P(x \geqq a+1,\ p) = \phi(\lambda) \quad . \tag{232}$$

Also, p^* and q^* should now be defined by $p^* = \frac{a+1}{n+1}$, $q^* = \frac{n-a}{n+1}$, and with this modification the formulae (226)-(231) become true even in this case.

Attention should be drawn to the fact that the equations presented so far are given in terms of n+1, not in terms of n. Also p^* is defined with n+1 in the denominator. In this way a sort of "automatic" continuity correction is made which replaces the classical half-unit correction.

3.2.4 <u>Application of the Beta-transformation to the confidence problem</u>: For the confidence interval problem, it is evident that the first transformation series in the form (226) provides the solution.

Let the desired two-sided confidence level be $1-2\varepsilon$. Let λ_ε be defined by

$$\phi(\lambda_\varepsilon) = 1-\varepsilon \quad . \tag{233}$$

For a given value of α we insert $\lambda = \lambda_\varepsilon$ and $p^* = \frac{a}{n+1}$ in the first terms of (226) and calculate p from the inverse of ϕ. According to the equation $p(x \geqq a,\ p) = 1-\phi(\lambda)$ and to (233), the value of p is an approximate solution of $p(x \geqq a,\ p) = \varepsilon$, and is consequently an approximation to the lower confidence limit p_ℓ. With $\lambda = -\lambda_\varepsilon$ and $p^* = (a+1)/n+1$, we obtain the upper confidence limit p_u by a similar procedure.

If we take the special value of α obtained by substituting $\alpha = \alpha_1(\lambda_\varepsilon)$ given by (222) in (226), we have what has been previously called the 'best' transformation. It follows that in this special case, the term of order $(n+1)^{-1}$ vanishes in (226) so that a comparatively simple formula is obtained.

From a practical point of view, the method outlined above is quite unimportant, as:

(i) excellent tables of confidence limits have been published, and

(ii) tables of percentage points of the incomplete Beta- and F-distributions can be used for this purpose.

Theoretically the method is, however, of a certain interest, and we shall therefore illustrate the dependence of the best transformation upon the prescribed confidence level by assigning some special values to the normal deviate in (233). In the series, the term of order $(n+1)^{-1}$ (the skewness correction) is also given, as it increases the applicability of the expansions to cases where this term differs from zero.

Case (i): $\lambda = 1, \quad \alpha_1(\lambda) = 0$

When λ is near 1 ($\varepsilon \sim 16\%$), it is best not to make any transformation at all, and (17.24) becomes

$$p = p^* - \lambda\sqrt{\frac{p^*q^*}{n+1}} + \frac{\lambda^2-1}{3}\,\frac{q^*-p^*}{n+1} + \ldots \quad . \tag{234}$$

Case (ii): $\lambda = 2, \quad \alpha_1(\lambda) = \tfrac{1}{2}$

When ε is near 2.5%, substituting α in (224) and integrating, we get the inverse sine transformation as the best transformation, and the formula (226) becomes:

$$2 \text{ arc } \sin\sqrt{p} = 2 \text{ arc } \sin\sqrt{p^*} - \lambda(n+1)^{-\frac{1}{2}} + \frac{\lambda^2-4}{12}(q^*-p^*)(p^*q^*)^{-\frac{1}{2}}(n+1)^{-1} + \ldots \quad , \tag{235}$$

or if the angle $\phi^* = \phi(p^*)$ is introduced in the last term (i.e., putting $p^* = \sin^2\frac{\phi^*}{2}$), we get

$$(q^*-p^*)(p^*q^*)^{-\frac{1}{2}} = \frac{\cos^2\frac{\phi^*}{2} - \sin^2\frac{\phi^*}{2}}{\sin\frac{\phi^*}{2}\cdot\cos\frac{\phi^*}{2}} = \frac{\cos\phi^*}{\frac{1}{2}\sin\phi^*} = 2\cot\phi^* \ .$$

In this case (235) becomes

$$\phi = \phi^* - \lambda(n+1)^{-\frac{1}{2}} + \frac{\lambda^2-4}{6}(n+1)^{-1}\cot\phi^* + \ldots \quad .$$

When 95% two-sided confidence intervals are required, it is thus advantageous to use the inverse sine transformation. This interesting property of the transformation shows that it may serve the double purpose of both stabilizing the variance and furnishing approximations to the tail of the Binomial distribution.

Case (iii): An interesting situation arises when λ is large. We should then theoretically (but not practically) apply the following transformation:

$$\phi(p, \tfrac{2}{3}) = \int_0^p \frac{dx}{x^{\frac{2}{3}}(1-x)^{\frac{2}{3}}} \quad , \tag{236}$$

where $\alpha = 1 - \frac{\lambda^2+2}{3\lambda^2} \to \frac{2}{3}$ as λ becomes large.

For the sake of comparison, two classical formulae will be mentioned here.

If a Binomial variable is expanded directly in a Cornish-Fisher series, and the expansion is solved with respect to p, the following equations are obtained (half-correction being used):

$$p_\ell = p^*-\tfrac{1}{2}n-\lambda_\varepsilon\sqrt{\frac{p^*q^*}{n}} + \frac{2\lambda_\varepsilon^2+1}{6n}(q^*-p^*) + \ldots$$

$$p_u = p^*+\tfrac{1}{2}n+\lambda_\varepsilon\sqrt{\frac{p^*q^*}{n}} + \frac{2\lambda_\varepsilon^2+1}{6n}(q^*-p^*) + \ldots \quad , \tag{237}$$

with $p^* = \frac{a}{n}$ in both formulae.

A related type of formulae is constructed by solving the equation:

$$a \pm \tfrac{1}{2} - np = \lambda_{\varepsilon}\sqrt{npq} \qquad (238)$$

with respect to p (cf. Cramer, 1946, p. 515).

In Table VIII a numerical comparison is made between (234), (235), (237) and the formula obtained by solving (238). The skewness correction has not been included in the calculations. The value p_{ℓ} (exact) has been obtained by interpolating in Thompson's (1941) tables.

It is seen from the table that the inverse sine formula gives excellent 95% values. This confirms the theory developed in this section. The 80 and 99% values are also good, and better than the values obtained by means of the other formulae.

The formula obtained by solving (238) can be ranked as second.

Equation (234) is rather inaccurate except in the 80% case, where it is fairly good (in accordance with the theory).

The classical formula (237) provides poor values. If it is used, the skewness correction (last term in the right member) ought to be included.

Table VIII. Lower confidence limits p_ℓ computed without skewness correction, n = 49.

confidence level two-sided	a	p_ℓ(exact)	p_ℓ computed from: (234)	(235)	(237)	(238)
	5	0.050	.046	.052	.036	.051
	10	0.131	.128	.133	.120	.132
80%	15	.220	.217	.221	.212	.220
	20	.312	.311	.313	.308	.313
	25	.410	.409	.410	.408	.410
	5	.034	.017	.033	.007	.038
	10	.102	.089	.102	.081	.107
95%	15	.182	.173	.182	.167	.187
	20	.270	.264	.270	.260	.273
	25	.363	.361	.363	.360	.365
	5	.023	< 0	.019	< 0	.030
	10	.080	.054	.077	.046	.089
99%	15	.152	.133	.150	.126	.161
	20	.234	.222	.232	.217	.241
	25	.323	.318	.322	.316	.327

3.2.5 <u>Application of the Beta-transformation to the Binomial tail problem</u>: Applying the Beta-transformation to the tail problem, we see that we should determine the normal deviate λ which corresponds to the tail of the Binomial, when λ has been calculated. We immediately obtain:

$$p(x \geq a, p) = 1 - \Phi(\lambda) \quad \text{or} \quad p(x \leq a, p) = \Phi(\lambda) \quad .$$

Let us apply here the second transformation series (228). (The application of (227) is similar.)

We shall assume that only the first term in the right member of (228) is computed (if higher accuracy is required the computations can be performed in two stages, the second term being included in the last stage where we substitute λ as computed in the first stage). Provided we know something in advance about the magnitude of λ, the equation (225) should, theoretically, govern our choice of α. If the same values are assigned to λ as in the confidence interval problem, we have the following transformations:

<u>Case (i)</u>: $\lambda = 1 \qquad \alpha_2(\lambda) = 1$

$$\phi(p^*,1) = \int_{p_0^*}^{p^*} x^{\alpha-1}(1-x)^{\alpha-1}dx = \int_{p_0^*}^{p^*} \frac{dx}{x(1-x)} = \log \frac{p^*}{q^*} - \log \frac{p_0^*}{q_0^*} \quad . \tag{239}$$

(The new integral is gotten from (224) by replacing p by p* and α by $1-\alpha$.)

Thus the logarithmic function which connects the Binomial distribution with the z-distribution is a prominent member of the class of Beta-transformation.

Case (ii): $\lambda = 2 \qquad \alpha_2(\lambda) = \frac{1}{2}$

$$\phi(p^*,\tfrac{1}{2}) = \int_{p_0^*}^{p^*} \frac{dx}{\sqrt{x(1-x)}} = 2 \text{ arc } \sin\sqrt{p^*} - 2 \text{ arc } \sin\sqrt{p_0^*} \; .$$

Hence the inverse sine transformation is the best and the first few terms of the expansion (228) run in this case as follows:

$$\lambda = (2 \text{ arc } \sin\sqrt{p^*} - 2 \text{ arc } \sin\sqrt{p})(n+1)^{\frac{1}{2}} + \frac{\lambda^2-4}{12}(q-p)(pq)^{-\frac{1}{2}}(n+1)^{-\frac{1}{2}} + \ldots \; . \tag{240}$$

This formula should be compared with a very similar expression which Freeman and Tukey (1950) have constructed empirically as we have seen previously. In their formula, the quantity $\frac{q-p}{\sqrt{(n+1)pq}}$ is replaced by $\frac{1}{\sqrt{np+1}} - \frac{1}{\sqrt{nq+1}}$.

Case (iii): $\lambda \to \infty \qquad \alpha_2(\lambda) = \frac{1}{3}$

This leads to the transformation

$$\phi(p^*,\tfrac{1}{3}) = \int_0^{p^*} \frac{dx}{x^{1/3}(1-x)^{1/3}} \; . \tag{241}$$

Again the inverse sine method is recommended here for practical use.

3.3 Transformation of the Poisson and χ^2 distributions:

3.3.1 Cornish-Fisher expansion and the best transformation of a Poisson variate: Regarding the Poisson as a limiting form of the Binomial case, we can treat the transformation problem of a Poisson variable in a similar way to the corresponding problem for a Binomial variate. In fact, we can utilize the results of the Binomial case as it stands in (213) and make the necessary changes which are imposed by the new characteristics of the quantities n and p. For that we use (213) after substituting a_1 and a_2 from (218) to get:

$$\phi(p)=\phi(p^*)-\lambda(p^*q^*)^{\frac{1}{2}}\phi'(p^*)(n+1)^{-\frac{1}{2}}+\tfrac{1}{2}\lambda^2[p^*q^*\phi''+\alpha_1(\lambda)(q^*-p^*)\phi'(p^*)] \cdot (n+1)^{-1}+\ldots \quad . \tag{242}$$

Now as n becomes very large and $np \to m$(constant), we find that $\phi(p)$ becomes $\phi(m)$, $np^* \to a$ or $p^* \to \frac{a}{n}$, so that $\phi(p^*)$ becomes $\phi(a)$; noting also that $\phi'(p)= \phi_m'(p)\frac{dm}{dp}= n\phi_m'(p)$, and $\phi''(p) = n^2\phi_m''(p)$, then (242) becomes

$$\phi(p)=\phi(p^*)-\lambda a^{\frac{1}{2}}\,\frac{(n-a+1)^{\frac{1}{2}}}{(n+1)^{\frac{3}{2}}}\,n\phi_m'(p^*)+\tfrac{1}{2}\lambda^2\Big[\frac{a(n-a+1)}{(n+1)^3}\phi_m''n^2 + \alpha_1\,\frac{n-2a+1}{(n+1)^2}\,n\phi_m'(p^*)\Big]+\ldots \quad .$$

Taking the limits as $n \to \infty$, we get

$$\phi(m) = \phi(a)-\lambda a^{\frac{1}{2}}\phi'(a)+\tfrac{1}{2}\lambda^2[a\phi''(a)+\alpha_1(\lambda)\phi(a)]+\ldots \quad . \tag{243}$$

This is the general transformation $\phi(x)$ of a Poisson variate X which we need for treating confidence and tail problems. Equating the expression within brackets in (243) to zero, replacing $\alpha_1(\lambda)$ by α and a by m, we have a simple differential equation which has the solution

$$\phi(m) = \int_{m_0}^{m} x^{-\alpha}\,dx = \begin{cases} \frac{1}{1-\alpha}\, m^{1-\alpha} + \text{const.} & \alpha < 1 \\ \log \frac{m}{m_0} & \alpha = 1 \end{cases} \quad . \tag{244}$$

Deriving ϕ', ϕ'' from (244) and substituting in (243) we get

$$\phi(m)=\phi(a)-\lambda a^{\frac{1}{2}-\alpha}+\frac{(2-3\alpha)\lambda^2-2}{6}a^{-\alpha}-\frac{1}{36}a^{-\frac{1}{2}-\alpha}[(1+6\alpha(\alpha-1))\lambda^3 + \lambda(12\alpha-7)]+\ldots \quad . \tag{245}$$

(244) and (245) may be considered as limiting expressions of (224) and (226). Analogously the second transformation series gives

$$\lambda = [\phi(a)-\phi(m)]m^{\alpha-\frac{1}{2}}+\frac{(3\alpha-1)\lambda^2-2}{6}m^{-\frac{1}{2}}-\frac{1}{72}[12\alpha^2-1)\lambda^3 - (24\alpha+2)\lambda]m^{-1}+\ldots \quad . \tag{246}$$

The discussion of the choice of α follows exactly the same lines as in the Binomial case. The results are summarized below.

3.3.2 <u>Determination of confidence limits for the mean of the Poisson distribution:</u> From (245) we obtain three confidence formulae by giving λ the values 1, 2, ∞. We then find:

<u>Case (i):</u> $\lambda = 1, \quad \alpha_1(\lambda) = 0$ with

$$m = a - \lambda a^{\frac{1}{2}} + \frac{\lambda^2 - 1}{3} - \frac{\lambda^3 - 7\lambda}{36} a^{-\frac{1}{2}} + \dots \quad ; \tag{247}$$

<u>Case (ii):</u> $\lambda = 2, \quad \alpha_1 = \frac{1}{2}$ with

$$m^{\frac{1}{2}} = a^{\frac{1}{2}} - \lambda + \frac{\lambda^2 - 4}{12} a^{-\frac{1}{2}} + \frac{\lambda^3 + 2\lambda}{72} a^{-1} \quad ; \tag{248}$$

<u>Case (iii):</u> $\lambda \to \infty, \quad \alpha_1(\lambda) \to \frac{2}{3}$ with

$$m^{\frac{1}{3}} = a^{\frac{1}{3}}[1 - \frac{\lambda}{3} a^{-\frac{1}{2}} - \frac{1}{9} a^{-1} + \frac{\lambda^3 - 3\lambda}{324} a^{-\frac{3}{2}} + \dots] \quad . \tag{249}$$

The lower confidence limit m_ℓ may be obtained from any of these formulae by substituting $\lambda = \lambda_\varepsilon$, and the upper limit m_u by substituting $\lambda = -\lambda_\varepsilon$ and replacing a by a+1.

3.3.3 <u>Determination of the tail of the Poisson distribution:</u>

The area of a tail of the Poisson distribution can, in analogy with the Binomial case, be calculated either from (245) or from (246) by assigning special values to α.

We shall list a few of these formulae in a form suitable for computation. If we omit the last term in (249) and solve with respect to λ, we have:

$$\lambda = 3\sqrt{a}\ (1 - (\frac{m}{a})^{\frac{1}{3}}) - \frac{1}{3\sqrt{a}} \quad . \tag{250}$$

If we put $\alpha = \frac{1}{3}$ in (246), we get a similar formula:

$$\lambda = \frac{3}{2}\sqrt{m}\ [(\frac{a}{m})^{\frac{2}{3}} - 1] - \frac{1}{3\sqrt{m}} \quad . \tag{251}$$

In Table IX a numerical comparison is made between (250), (251) and the formula obtained by taking $\alpha = \frac{1}{2}$ in (246). This formula may be used in two ways:

(a) $\lambda = 2\sqrt{a} - 2\sqrt{m}$.

(b) $\lambda = \lambda_0$ is computed from (252), and the value then obtained is inserted in (252)

$$\lambda = \lambda_0 + \frac{\lambda_0^2 - 4}{12}\frac{1}{\sqrt{m}} \quad . \tag{253}$$

In Table IX, the column denoted by λ(exact) gives the correct value of λ computed by means of Molina's (1947) table. Apart from the fact that (253) gives poor values when λ is very different from $\lambda = 2$, all the formulae are fairly accurate.

3.3.4 <u>Determination of percentage points of the χ^2-distribution</u>: Let X be a Poisson variate with mean m so that

Table IX. Numerical study of the accuracy of formulae (250) - (253).

mean	a	λ(exact)	λ computed from (17.49)	(17.50)	(17.51)	(17.52)
m = 1	1	- .34	- .33	- .33	.00	- .33
	2	.63	.64	.55	.83	.55
	3	1.40	1.40	1.29	1.46	1.31
	4	2.08	2.05	1.95	2.00	2.00
	5	2.68	2.64	2.55	2.47	2.65
	6	3.24	3.17	3.12	2.90	3.27
	7	3.77	3.66	3.66	3.29	3.86
	8	4.27	4.12	4.17	3.66	4.44
m = 2	1	-3.92	-3.80	-3.83	-4.32	-3.94
	2	-3.29	-3.25	-3.23	-3.50	-3.28
	3	-2.77	-2.76	-2.72	-2.86	-2.75
	4	-2.31	-2.31	-2.27	-2.32	-2.29
	5	-1.89	-1.89	-1.86	-1.85	-1.87
	10	- .11	- .11	- .11	.00	- .11
	15	1.38	1.38	1.37	1.42	1.37
	16	1.66	1.66	1.64	1.68	1.64
	17	1.93	1.92	1.91	1.92	1.91
	18	2.19	2.19	2.71	2.16	2.18
	20	2.70	2.69	2.68	2.62	2.70
	25	3.90	3.88	3.89	3.68	3.93
	26	4.13	4.11	4.12	3.87	4.16
	27	4.36	4.33	4.35	4.07	4.40

$$\Pr(x \geq a;\ m) = \int_0^1 \frac{m^a}{\Gamma(a)}(1-t)^{a-1}\ e^{-m(1-t)}\ dt \quad ,$$

and putting $m(1-t) = \frac{1}{2}\chi^2$, we get

$$\Pr(x \geq a;\ m) = \frac{1}{2^a\Gamma(a)} \int_0^{2m} e^{-\frac{1}{2}\chi^2}\ (\chi^2)^{a-1}\ d\chi^2$$

$$= \Pr(\chi^2_{2a} \leq 2m) \quad . \tag{254}$$

Let us denote the degrees of freedom of the χ^2-variate by f. Replacing a by $\frac{1}{2}f$, m by $\frac{1}{2}\chi^2$, and λ by $(-\lambda)$ in (251) and (245) and remembering that the best transformation is $\phi(\frac{1}{2}\chi^2) = \frac{1}{1-\alpha}(\frac{1}{2}\chi^2)$ we get for $\alpha < 1$:

$$\frac{1}{1-\alpha}(\tfrac{1}{2}\chi^2)^{1-\alpha} = \frac{1}{1-\alpha}(\tfrac{1}{2}f)^{1-\alpha} + \lambda(\tfrac{1}{2}f)^{\frac{1}{2}-\alpha} + \frac{(2-3\alpha)\lambda^2-2}{6}(\tfrac{1}{2}f)^{-\alpha}$$

$$+ \frac{1}{36}(\tfrac{1}{2}f)^{-\frac{1}{2}-\alpha}[(1+6\alpha(\alpha-1))\lambda^3+(12\alpha-7)\lambda] \quad .$$

Multiplying both sides by $(\frac{1}{2})^{\alpha-1}$, we get:

$$\frac{1}{1-\alpha}(\chi^2)^{1-\alpha} = \frac{1}{1-\alpha}\ f^{1-\alpha} + \lambda\sqrt{2}\ f^{\frac{1}{2}-\alpha} + \frac{(2-3\alpha)\lambda^2-2}{3}\ f^{-\alpha}$$
$$+ \frac{\sqrt{2}}{18}\ [(1+6\alpha(\alpha-1))\lambda^3+(12\alpha-7)\lambda]f^{-\frac{1}{2}-\alpha} \ldots \quad . \tag{255}$$

For $\alpha = 0$ we have a series given by Peiser (1943) and Goldberg & Levine (1946),

$$\chi^2 = f + \lambda\sqrt{2f} + \frac{2(\lambda^2-1)}{3} + \frac{\sqrt{2}(\lambda^3-7\lambda)}{18\sqrt{f}} + \ldots \quad . \tag{256}$$

For $\alpha = \frac{1}{2}$ we have:

$$\sqrt{2\chi^2} = \sqrt{2f} + \lambda + \frac{\lambda^2-4}{6}\frac{1}{\sqrt{2f}} - \frac{\lambda^3+2\lambda}{18}\frac{1}{2f} + \ldots \quad . \tag{257}$$

It is of interest to compare this formula with Fisher's χ^2-approximation

$$\sqrt{2\chi^2} \sim \sqrt{2f-1} + \lambda \quad .$$

Finally for $\alpha = \frac{2}{3}$ we obtain after dividing by $3f^{\frac{1}{3}}$ and putting $c = \frac{2}{9f}$:

$$\left(\frac{\chi^2}{f}\right)^{\frac{1}{3}} = 1 + \lambda c^{\frac{1}{2}} - c - \frac{\lambda^3-3\lambda}{} c^{\frac{3}{2}} + \ldots \quad . \tag{258}$$

If the last term is neglected, we have Wilson and Hilferty's χ^2-approximation. Thus we conclude that the results of these authors may be regarded as variations of one single theme. The three formulae are accurate, if several terms are computed, but if only the first two terms are retained, the desired confidence level should determine which formula is to be preferred.

We finally observe that the logarithmic transformation obtained in (244) for the case $\alpha = 1$ has been applied to the χ^2-distribution by Bartlett and Kendall (1946).

3.4 Negative Binomial: The theory which has been developed in the preceding cases can be extended to the case of a negative Binomial variable. Only a few of the results will be briefly mentioned here.

We denote the parameters of the negative Binomial variable x by N, P and Q = 1+P. The probability that x assumes a given value A is given by:

$$\Pr(x = A) = \binom{N+A-1}{A} P^A Q^{-N-A} \quad .$$

We introduce the notation $P^* = \frac{A}{N}$. The analogues of (213), (214) are of the form

$$\Psi(P) = \Psi(P^*) + A_1 N^{-\frac{1}{2}} + A_2 N^{-1} + \ldots \quad , \tag{259}$$

where the coefficients A_ν depend upon P*, but not upon P, and:

$$\Psi(P^*) = \Psi(P) + B_1 N^{-\frac{1}{2}} + B_2 N^{-1} + \ldots \quad , \tag{260}$$

where the coefficients B_ν depend upon P, but not upon P*.

The coefficients of N^{-1} in both expansions vanish if $\Psi(P)$ is of the form

$$\Psi(P;\ \alpha,\beta) = \int_{P_0}^{P} x^{-\alpha}(1+x)^{-\beta}\, dx \quad . \tag{261}$$

In (259) the parameters α and β should be chosen as follows:

$$\alpha = 1 - \frac{\lambda^2+2}{3\lambda^2} \, , \qquad \beta = \frac{2(\lambda^2+2)}{3\lambda^2} \quad . \tag{262}$$

In (260) the parameter values are:

$$\alpha = \frac{\lambda^2+2}{3\lambda^2}, \qquad \beta = 1 - \frac{2(\lambda^2+2)}{3\lambda^2} \quad . \tag{263}$$

By assigning special values to λ, a variety of different transformations is obtained. We give here only two brief remarks:

(i) If we take $\lambda = 2$ in (263), we find $\alpha = \frac{1}{2}$, $\beta = 0$, and $\Psi(P) = 2P^{\frac{1}{2}}$. Thus, when the tail of the negative Binomial distribution corresponds to a normal deviate equal to about 2, a simple square root transformation is preferred to an inverse sinh transformation.

(ii) The inverse sinh transformation is not 'best' for any value of λ in the meaning which we have assigned to the term here. It belongs, however, to the class of functions (261) and is obtained for $\alpha = \beta = \frac{1}{2}$.

4. Z-transformation.

4.1 Properties of the Z-transformation: With large samples and moderate or small correlations, the correlation obtained from a sample of n pairs of values is distributed normally about the true value ρ, with variance $= \frac{(1-\rho^2)^2}{n-1}$. It is therefore usual to attach to an observed value r a standard error $\frac{1-r^2}{\sqrt{n-1}}$ or $\frac{1-r^2}{\sqrt{n}}$. This procedure is only valid under the restrictions stated above. With small samples the value of r is often very different from the true value ρ and the factor $1-r^2$ correspondingly in error. In addition, the distribution of r which can be written as:

$$f(r,\rho) = \frac{(n-2)(1-\rho^2)^{\frac{1}{2}(n-1)}}{\pi}(1-r^2)^{\frac{1}{2}(n-4)}\int_0^{\infty}\frac{dw}{(\cosh w-r\rho)^{n-1}}, \tag{264}$$

$-1 \leq r \leq +1$, is far from normal, so that tests of significance based on the large sample formula are often very deceptive. Since it is with small samples less than 100, that the practical research worker ordinarily wishes to use the correlation coefficient, Fisher introduced the transformation

$$Z = \tfrac{1}{2}\{\log_e(1+r)-\log_e(1-r)\} = r + \tfrac{1}{3}r^3 + \tfrac{1}{5}r^5 \ \ldots, \tag{265}$$

which leads approximately to the normal distribution. We notice that as r changes from 0 to 1 Z will pass from 0 to ∞.

For small values of r, Z = r approximately, but as r approaches unity, Z increases without limit. For negative values of r, Z is negative. The advantage of this transformation of r into Z lies in the distribution of these two quantities in random samples. The standard deviation of r depends on the true value of the correlation ρ as is seen from the formula $\sigma_r = \frac{1-\rho^2}{\sqrt{n-1}}$. Since ρ is unknown, we have to substitute for it the observed value r and this value will not in small samples be a very accurate estimate of ρ. The mean of Z is $\delta = \frac{1}{2}\log_e \frac{1+\rho}{1-\rho}$ or slightly more accurately $\frac{1}{2}\log_e \frac{1+\rho}{1-\rho} + \frac{\rho}{2(n-1)}$ + terms in $\frac{1}{(n-1)^2}$. The standard error of Z is $(n-3)^{-\frac{1}{2}}$, practically independent of the value of the correlation in the population from which the sample is drawn.

In the second place the distribution of r is not normal in small samples and even for large samples it remains far from normal for high correlations. The distribution of Z is not strictly normal, but it tends to normality rapidly as the sample is increased, whatever may be the value of the correlation. In fact it can be shown that for the distribution of Z,

$$\beta_1 = \rho^6(n-1)^{-3} + 0((n-1)^{-2}) \quad ,$$
$$\beta_2 = 3 + 2(n-1)^{-1} + 0((n-1)^{-2}) \quad . \tag{266}$$

Thus for $n = 11$, β_1 is of the order of .001 even if ρ is high and β_2-3 is of the order 0.2. In such a case the true mean of Z differs from $\delta = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho} = \text{arctanh}\ \rho$ by 0.05 which is not large but might be important in some cases.

Finally the distribution of r changes its form rapidly as ρ is changed. On the contrary the distribution of Z is nearly insensitive to changes in ρ.

4.2 <u>Applications of the Z-transformation</u>: We can utilize this transformation in the following situations:

1. To test the hypothesis $\rho = \rho^*$ (ρ^* known). For we consider the variate

$$u = (\text{arctanh}\ r - \text{arctanh}\ \rho^*)\sqrt{n-3}$$

as approximately normally distributed $N(0,1)$.

2. Suppose that K-bivariate normal populations have correlation coefficients $\rho_1, \rho_2 \ldots \rho_k$ and that $r_1, r_2 \ldots r_k$ are the estimates based on samples of size $n_1, n_2 \ldots n_k$, respectively. To test the hypothesis H_o: $\rho_1=\rho_2=\ldots=\rho_k=\rho^*$ (ρ^* known), the quantity $U=\sum_{i=1}^{k}(\text{arctanh}\ r_i-\text{arctanh}\ \rho^*)^2(n_i-3)$

has an approximate χ^2 distribution with K degrees of freedom. We reject H_o with probability α if $U \geq \chi^2_\alpha(k)$, where $\chi^2_\alpha(k)$ is the appropriate χ^2 value for K degrees of freedom and probability α.

3. Suppose in the preceding case we wish to test H_o: $\rho_1=\rho_2=\ldots=\rho_k$ without specifying the common value of the ρ's. Then we can use $W = \Sigma(n_i-3)(Z_i-\overline{Z})^2$ as a χ^2 variate with K-1 d.f. where:

$$Z_i = \text{arctanh } r_i \quad \text{and} \quad \overline{Z} = \frac{\Sigma(n_i-3)Z_i}{\Sigma(n_i-3)} \quad .$$

4. If $\rho_1=\rho_2=\ldots=\rho_k$ in the preceding case, then the "best" linear combined estimate r of the common correlation ρ is given by $r = \tanh Z^*$ where $Z^* = \overline{Z} - \frac{m\rho^*}{2}$ and $\overline{Z}$ is defined in (3) and $\rho = \frac{1}{K}\Sigma r_i$, $m = \frac{\Sigma[(n_i-3)/(n_i-1)]}{\Sigma(n_i-3)}$.

5. To set approximate $1-\alpha$ confidence limits on ρ. Here we use the fact that $(n-3)^{\frac{1}{2}}(\text{arctanh } r - \text{arctanh } \rho)$ is approximately distributed $N(0,1)$. A confidence interval at the level $1-\alpha$ on ρ is:

$$\tanh(\text{arctanh } r - \frac{Z_{\alpha/2}}{\sqrt{n-3}}) \leq \rho \leq \tanh(\text{arctanh } r + \frac{Z_{\alpha/2}}{\sqrt{n-3}}) \quad ,$$

where $Z_{\alpha/2}$ is the appropriate value for the standardized normal distribution.

Finally, R. A. Fisher has tabulated $\frac{1}{2}\log_e \frac{1+x}{1-x}$ for various values of x and many of the approximate tests can be readily applied by using his tables. However if $n \leq 25$, one should use the tables prepared by David (1938) who has tabulated the integral

$$\int_{-1}^{\rho_0} f(r,\rho)dr = Pr(r \leq \rho_0/\rho) \tag{267}$$

for various values of ρ_0, ρ and n, where $f(r,\rho)$ is the density function of r mentioned in (264). If $n > 25$, then the Z-transformation is satisfactory.

5. Transformations of χ^2 towards normality.

5.1 Some properties of χ^2-distribution: The density of χ^2 is

$$dF = \frac{1}{2^{\frac{1}{2}\nu}\,\Gamma(\frac{1}{2}\nu)} e^{-\frac{1}{2}\chi^2} (\chi^2)^{\frac{1}{2}\nu-1} d(\chi^2) \qquad 0 \leqq \chi^2 \leqq \infty . \tag{268}$$

The characteristic function is:

$$\phi(t) = \frac{1}{(1-2it)^{\nu/2}} . \tag{269}$$

Whence for the cumulants we have

$$K_r = \nu \cdot 2^{r-1} (r-1)! . \tag{270}$$

The central moments are

$$\begin{aligned} \mu_2 &= 2\nu \\ \mu_3 &= 8\nu \\ \mu_4 &= 48\nu + 12\nu^2 . \end{aligned} \tag{271}$$

As $\nu \to \infty$ the χ^2-distribution tends to normality, for in standard measure we have

$$\phi(t) = e^{-\frac{\nu it}{\sqrt{2\nu}}} \left(1 - \frac{2it}{\sqrt{2\nu}}\right)^{-\nu/2} , \tag{272}$$

so that

$$\log \phi \to \frac{-\nu it}{\sqrt{2\nu}} - \frac{\nu}{2}\left[\frac{-2it}{\sqrt{2\nu}} - \tfrac{1}{2}\left(\frac{2it}{\sqrt{2\nu}}\right)^2 \ldots\right] \to -\tfrac{1}{2}t^2$$

as $\nu \to \infty$. The tendency is rather slow because the first term ignored is of order $\nu^{-\frac{1}{2}}$. The skewness and kurtosis coefficients are

$$\gamma_1 = (8/\nu)^{\frac{1}{2}}$$
$$\gamma_2 = \frac{12}{\nu} \quad . \tag{273}$$

5.2 <u>The transformation $\sqrt{2\chi^2}$</u> : This transformation was suggested by Fisher for approaching normality more rapidly than χ^2. For the variate χ the *r*th moment is

$$\mu_r' = 2^{r/2} \frac{\Gamma(\frac{\nu+r}{2})}{\Gamma(\frac{\nu}{2})} \quad . \tag{274}$$

Using Stirling's expansion

$$\log \Gamma(x+1) \sim \tfrac{1}{2} \log(2\pi) + (x+\tfrac{1}{2}) \log x - x + \frac{1}{12x} - \frac{1}{360x^3} + \ldots ,$$

we find

$$\mu_1' = \sqrt{\nu} \left(1 - \frac{1}{4\nu} + \frac{1}{32\nu^2} + \frac{5}{128\nu^3} + \ldots\right) ,$$

whence

$$\mu_1'^2 = \nu\left(1 - \frac{1}{2\nu} + \frac{1}{8\nu^2} + \frac{1}{16\nu^3} + \ldots\right) .$$

Moreover $\mu_2' = \nu$, $\mu_3' = (\nu+1)\mu_1'$, $\mu_4' = (\nu+2)\nu$. Whence for the central moments

$$\mu_2 = \frac{1}{2} - \frac{1}{8\nu}$$
$$\mu_3 = \frac{1}{4\sqrt{\nu}} + \ldots \tag{275}$$
$$\mu_4 = \frac{3}{4} - \frac{3}{8\nu} + 0(\nu^{-2}) \quad .$$

Hence for $\sqrt{2\chi^2}$ we have

$$\mu_2 = 1 - \frac{1}{4\nu} + \ldots$$

$$\gamma_1 = \frac{1}{\sqrt{2\nu}} + \ldots \tag{276}$$

$$\gamma_2 = 0(\nu^{-2}) \quad .$$

A comparison with (273) shows that $\sqrt{2\chi^2}$ tends to normality with considerably greater rapidity than χ^2. Moreover, the expression for μ_1' for χ can be written as $\mu_1' = \sqrt{\nu}(1-\frac{1}{2\nu})^{\frac{1}{2}} = \sqrt{\nu-\frac{1}{2}}$ to order $\nu^{-\frac{3}{2}}$ and hence $\sqrt{2\chi^2}$ is distributed about mean $\sqrt{2\nu-1}$, to the same order, with variance which is unity to order ν^{-1}.

5.3 <u>The transformation $(\chi^2/\nu)^{\frac{1}{3}}$</u>: This transformation was first found by Wilson-Hilferty (1931) as an improvement on the preceding Fisher transformation. Let us find the distribution of $(\frac{\chi^2}{\nu})^h = y$. Writing $\xi = \chi^2 - \nu$, we have

$$y = (1 + \frac{\xi}{\nu})^h = 1 + \frac{h\xi}{\nu} + \frac{h(h-1)}{2!}\frac{\xi^2}{\nu^2} + \ldots \quad . \tag{277}$$

Taking mean values and using the results for the moments of χ^2 which represent the raw moments of ξ, we find

$$\mu_1(y) = 1 + \frac{h(h-1)}{2!}\frac{\mu_2(\chi^2)}{\nu^2} + \ldots = 1 + \frac{h(h-1)}{\nu} + \frac{h(h-1)(h-2)(3h-1)}{6\nu^2} + \frac{h^2(h-1)^2(h-2)(h-3)}{6\nu^3} + 0(\nu^{-4}) \quad . \tag{278}$$

If we put rh = h we find the mean value of y^r and thus:

$$\mu_r'(y) = 1 + \frac{rh(rh-1)}{\nu} + \ldots \quad , \quad \text{hence}$$

$$\mu_2'(y) = 1 + \frac{2h(2h-1)}{\nu} + \frac{2h(2h-1)(2h-2)(6h-1)}{6\nu^2} + \ldots$$

$$\mu_3'(y) = 1 + \frac{3h(3h-1)}{\nu} + \frac{3h(3h-1)(3h-2)(9h-1)}{6\nu^2} + \ldots$$

$$\mu_4'(y) = 1 + \frac{4h(4h-1)}{\nu} + \frac{4h(4h-1)(4h-2)(12h-1)}{6\nu^2} + \ldots$$

$$\begin{aligned}\mu_3(y) &= \mu_3'(y) - 3\mu\mu_2' + 2\mu^3 \\ &= \frac{1}{6\nu^2}[3h(3h-1)(3h-2)(9h-1)-3h(h-1)(h-2)(3h-1) \\ &\quad -6h(2h-1)(2h-2)(6h-1)-36h^2(h-1)(2h-1) \\ &\quad +6h(h-1)(h-2)(3h-1)+36h^2(h-1)^2] \\ &\quad + \frac{1}{6\nu^3}[8h^2(3h-1)^2(3h-2)(2h-3)-3h^2(h-1)^2(h-2)(h-3) \\ &\quad -12h^2(2h-1)^2(2h-2)(2h-3)-6h^2(2h-1)(h-1)(h-2)(3h-1) \\ &\quad -6h^2(h-1)(2h-1)(2h-2)(6h-1)+12h^3(h-1)^3 \\ &\quad +6h^2(h-1)^2(h-2)(h-3)+12h^2(h-1)^2(h-2)(3h-1)] \\ &\quad + O(\nu^{-4}) \quad .\end{aligned}$$

If we now take 3h = 1, then the term in ν^{-2} vanishes.

Consequently,

$$\mu_1'(y) = 1 - \frac{2}{9\nu} + \frac{80}{3^7\nu^3} + O(\nu^{-4}) \quad .$$

$$\mu_2'(y) = 1 - \frac{2}{9\nu} + \frac{4}{3^4\nu^2} + \frac{56}{3^7\nu^3} + O(\nu^{-4}) \quad .$$

$$\mu_3'(y) = 1 \quad .$$

$$\mu_4'(y) = 1 + \frac{4}{9\nu} - \frac{4}{3^3\nu^2} + \frac{80}{3^7\nu^3} + 0(\nu^{-4}) \quad .$$

The central moments $\mu_2(y)$, $\mu_3(y)$, $\mu_4(y)$ are now:

$$\mu_2(y) = \frac{2}{9\nu} - \frac{104}{3^7\nu^3} + 0(\nu^{-4}) \quad .$$

$$\mu_3(y) = \frac{32}{3^6\nu^3} + 0(\nu^{-4}) \quad . \tag{279}$$

$$\mu_4(y) = \frac{4}{3^3\nu^2} - \frac{16}{3^6\nu^3} + 0(\nu^{-4}) \quad .$$

For the measures of skewness we have

$$\gamma_1 = \frac{2^{\frac{7}{2}}}{3^3\nu^{\frac{3}{2}}}$$

$$\gamma_2 = \frac{-4}{9\nu} \quad . \tag{280}$$

Comparison with (273), (276) shows that $(\frac{\chi^2}{\nu})^{\frac{1}{3}}$ tends to symmetry (as measured by γ_1) more rapidly than either χ^2 or $\sqrt{2\chi^2}$. By (279) the variance equals $\frac{2}{9\nu}$ to order ν^{-2} and the mean $1 - \frac{2}{9\nu}$. The result may be expressed by saying that $\{(\frac{\chi^2}{\nu})^{\frac{1}{3}} + \frac{2}{9\nu} - 1\}(\frac{9\nu}{2})^{\frac{1}{2}}$ is $N(0,1)$, approximately.

6. <u>Transformation of discrete variables by using an auxiliary variable</u>:

Eudy (1949) and Stevens (1950), separately, suggested the idea of obtaining exact confidence intervals for the parameter of a discrete distribution by adding a continuous random variable distributed rectangularly [0,1] to each discrete observation. In (1954) David investigated the use of an auxiliary variable in the transformation of Poisson, Binomial and Negative Binomial variates. He examined the following transformations:

(a) $$T = (x + z)^{\frac{1}{2}} \quad ,$$

where x is a Poisson variate with parameter λ and z is rectangular [0,1]. The transformed variable appears to be reasonably fitted by a normal curve, and it would seem that an assumption of normality to be used for the derivation of confidence intervals and testing hypothesis would not lead to serious risk of error, even for λ as small as 0.5.

(b) $$T = \sqrt{n+\tfrac{1}{2}}\ \sin^{-1}\sqrt{\frac{x+z}{n+1}} \quad ,$$

where x is a Binomial variable with index n and mean np, and z is rectangular [0,1].

Treating T as a normal variate, we can compute the tail probability of the variate x by use of normal tables. Moreover, David reports that this transformation is as good as that suggested by Anscombe $(T_A = \sqrt{n+\frac{1}{2}}\, \sin^{-1}\sqrt{\frac{x+3/8}{n+3/4}})$ for stabilizing the Binomial variance.

(c) $$T = \sinh^{-1} (x+z)^{\frac{1}{2}} \quad ,$$

where x is a Negative Binomial with index K and mean p, and z is rectangularly distributed [0,1] as before.

The variate T turns out to be nearly normally distributed, but the variance is not stable under changes of K and p.

Finally, for the case of a Poisson variate, one should refer to the work of Taylor and Tweedie (1955), who give explicit formulae for the mean, variance, γ_1 and γ_2 of the transformed variate in each of the three cases:

1. The auxiliary variable z is constant.
2. The auxiliary variable z is continuous uniform.
3. The auxiliary variable z is discrete uniform.

7. The Probit transformation.

7.1 Introduction: In the majority of biological assays there are two things to be considered, the stimulus (vitamine, drug, physical force, etc.) and the subject (an animal, a plant, a piece of tissue, etc.). The stimulus is applied to the subject at an intensity specified in units of concentration, weight, time, or other appropriate measure and as a result of which a response is produced by the subject. Different stimuli are then compared in terms of the magnitudes of the responses they produce, or more commonly in terms of the intensities required to produce equal responses. When the characteristic response is quantal, for any one subject, under controlled conditions, there will be a certain level of intensity below which the response does not occur and above which the response occurs; such a value is called the tolerance value. It varies from one member of the population to another, frequently between quite wide limits. When the characteristic response is quantitative, we have similar variation between individuals.

7.2 The frequency distribution of a tolerance: If the dose is measured by λ, the distribution of tolerances may be expressed by

$$dP = f(\lambda)d\lambda \quad . \tag{281}$$

This equation states that a proportion, dP, of the whole population consists of individuals whose tolerances lie between λ and $\lambda + d\lambda$. If a dose λ_0 is given to the whole population, all individuals will respond whose tolerances are less than λ_0 and the proportion of these is

$$P = \int_0^{\lambda_0} f(\lambda)d\lambda \quad . \tag{282}$$

The measure of dose ranges from zero to ∞, response being certain for very high doses, so that

$$\int_0^{\infty} f(\lambda)d\lambda = 1 \quad .$$

The distribution of tolerance, as measured on the natural scale, may be markedly skew on account of a few subjects with extremely high tolerances. Normalization can often be effected by expressing the tolerances in terms of the logarithm of the concentrations. Thus for much insecticidal work, if λ is the concentration of the toxic agent we take as the variate under consideration the quantity $x = \log_{10} \lambda$. The word dose is usually restricted to the scale of λ, and x is referred to as the measure of dosage. A graph of the percentage responding against a dose will give a steadily rising curve. The rate of increase in response per unit increase in dose is frequently very low in the region of zero or 100% response, but higher in the intermediate region, so that the curve is sigmoidal. When the stimulus is measured in dosage units, the curve takes the characteristic normal sigmoid form. It should be noted that if an insect is exposed to a dose, the probability that it will respond is P; hence if a batch of n insects is exposed to the dose λ_0 and all react independently, then

$$\Pr(r \text{ "responders"}) = \frac{n!}{r!(n-r)!} P^r Q^{n-r} \quad . \tag{283}$$

Finally, we define the median effective dose (referred to as E.D. 50, or L.D. 50) as the value Λ of λ for which

$$\int_0^{\Lambda} f(t)dt = 0.5 \quad . \tag{284}$$

When we transform to $x = \log_{10} \lambda$, the equation (281) becomes

$$dP = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \quad . \tag{285}$$

Thus μ is the population value of the mean dosage tolerance and efforts must be directed towards estimating it ($\mu = \log$ E.D. 50).

7.3 <u>The Probit transformation:</u> The probability integral of the proportion p is defined as the abscissa which corresponds to a probability P in a normal distribution N(5,1). In symbols the probit of p is Y, where:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y-5} e^{-\frac{1}{2}u^2} \, du \quad . \tag{286}$$

The effect of transforming from proportions to probits is illustrated in Fig. XI. Along the left-hand vertical axis is a linear scale of percentages with their corresponding probit values, and on the right-hand axis is a linear scale of probits with their corresponding percentage values. The transformation may be considered as a stretching of the left-hand scale to give that on the right-hand, during which process the sigmoid curve becomes a straight line. We know that the expected proportion of insects killed by a dosage x_0 is

$$P = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x_0} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \, dx \quad . \tag{287}$$

Comparing this with (286) we find that the probit of the expected proportion killed is related to the dosage by the linear equation

$$Y = 5 + \frac{1}{\sigma}(x-\mu) \quad . \tag{288}$$

By means of the probit transformation, experimental results may be used to give estimates of σ and μ. In particular, the median effective dosage is estimated as that value of x which corresponds to Y = 5.

7.4 <u>The Probit regression line</u>: When experimental data on the relationship between dose and mortality has been obtained, either a graphical or arithmetical process can be used to estimate the parameters. In order to make either type of estimate, the percentage kill observed for each dose must first be calculated and converted to probits. The probits are then plotted against x, the logarithm of the dose, and a straight line drawn by eye to fit the points as satisfactorily as possible. This line may be used to initiate the arithmetical process of estimating a better fitting line. The empirical probits plotted for a carefully conducted experiment often lie close to a straight line. The logarithm of L.D. 50 is estimated from the line as m, the dosage at which Y = 5. The slope of the line, b, which is an estimate of $1/\sigma$, is obtained as the increase in Y for a unit increase in x. These two parameters are then substituted in (288) to give an estimated relationship between dosage and kill.

Here the variance of the proportion killed is $\frac{PQ}{n}$ and this is inversely proportional to n. The reciprocal $\frac{n}{PQ}$ represents the weight to be attached to the observation on the batch in respect of the information it provides on P. We shall see that the weight to be attached to the probit of P is nw, where $w = \frac{Z^2}{PQ}$. Here Z is the ordinate of the normal distribution corresponding to the probability p, and may be written $Z = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(Y-5)^2}$.

7.5 Working probits and the arithmetical methods: Suppose now out of n_t individuals who are given the stimulus x_t (t=1,2...K), r_t respond. We may write $P_t = 1-q_t = \frac{r_t}{n_t}$, and the corresponding probits y_t will be obtained from the tables as indicated in Section (7.4). We fit the regression line y = a+bx to the data obtained (namely the paired values (y_t, x_t)).

Writing $n_t w_t$ for the weight of y_t and omitting the subscript t for convenience, standard regression theory provides the solution as follows:

$$b = \text{estimate of } \frac{1}{\sigma} = \frac{\Sigma nw(x-\bar{x})(y-\bar{y})}{\Sigma nw(x-\bar{x})^2} = \frac{\Sigma nw \Sigma nwxy - \Sigma nwx \Sigma nwy}{\Sigma nw \Sigma nwx^2 - (\Sigma nwx)^2},$$

$$a = \text{estimate of } (5 - \frac{\mu}{\sigma}) = \bar{y} - b\bar{x} \quad , \tag{289}$$

where $\bar{x} = \frac{\Sigma nwx}{\Sigma nw}$, $\bar{y} = \frac{\Sigma nwy}{\Sigma nw}$.

The maximum likelihood method of estimation leads to equations (289) in which the probits y_t are replaced by the working probits defined in (290), and these are used in conjunction with weights w_t given by $w_t = \frac{Z_t^2}{P_t Q_t}$, P, Q, and Z being as defined in Sections (7.3) and (7.4). Since the working probits and weights can only be estimated from the data, the solution must be found by iteration. This may be carried out as follows:

(a) From the observed p_t, find y_t from tables (Table 6 in Biometrika tables (1958) is recommended) and make a rough plot of y_t against x_t.

(b) As a first approximation a straight line fitted by eye to the points (y_t, x_t) may be used to obtain a set of provisional or expected probits Y_t' corresponding to x_t.

(c) The next step is to obtain a series of working probits y_t' and their weighting coefficients w_t' to use in relations (289) and so obtain a second approximation to the regression line. These working probits are obtained from either

$$y = Y + \frac{Q}{Z} - \frac{q}{Z} \, , \tag{290}$$

or

$$y = Y - \frac{P}{Z} + \frac{p}{Z} \quad , \tag{291}$$

as most convenient, where we find the maximum working probit $Y_{max} = Y + \frac{Q}{Z}$ or the minimum working probits $Y_{min} = Y - \frac{P}{Z}$ and the range $Y_{max} - Y_{min} = \frac{1}{Z}$, by entering Table 6 in Biometrika Tables (1958) with the expected probits Y_t' .

(d) The appropriate weights $n_t \, w_t'$ are found, entering the table with Y_t' .

(e) These weights and the working probits y_t' are then used to compute the coefficients a and b of equations (289).

(f) The values of the probits y_t'' obtained from this second approximation to the regression line may be used to repeat the cycle of the iterative process, the table being entered to obtain fresh working probits y_t'' and weights $n_t \, w_t''$. The iteration is continued until the process has 'settled down'.

8. The logistic function and its application to Bio-assay data:

8.1 Introduction and notation:

Berkson (1944), following the method used in probit transformations, has utilized the function $Q = \frac{1}{1+e^{\alpha-\beta x}}$ calling it the logistic function. The approximate shape of this function is close to that of the normal distribution function, and may have a better theoretic basis than the integrated normal curve. In the logistic function there are two parameters α and β, which, if known, determine the effect of any dose. The L.D. 50 is given by $\frac{\alpha}{\beta}$. The following definitions are used in this section:

Q = "true" logistic function of mortality rate for a given x, with parameters α, β.

$P = 1 - Q$.

$\hat{Q}$ = fitted function corresponding to estimates a and b of the parameters.

x_i = dose in log-units.

q_i = observed mortality rate for dose $x_i = \frac{m_i}{n_i}$.

$p_i = 1 - q_i$.

m_i = number of dead at x_i.

n_i = number exposed at x_i.

$\lambda_{50} = \frac{\alpha}{\beta}$ = log L.D. 50, the dose for which the mortality rate is 50%.

$$\ell = \ell n \frac{p_i}{q_i} .$$

L = logistic function in terms of logits corresponding to Q.

$\hat{L}$ = logistic function in terms of logits corresponding to $\hat{Q}$.

$$Q = \frac{1}{1 + e^{\alpha-\beta x}} . \qquad (292)$$

$$\hat{Q} = \frac{1}{1 + e^{a-bx}} . \qquad (293)$$

$$\ell n \frac{1-\hat{Q}}{\hat{Q}} = \hat{L} = a - bx . \qquad (294)$$

$$\frac{\partial \hat{Q}}{\partial a} = - \hat{Q}(1-\hat{Q}) = - \hat{Q}\hat{P} . \qquad (295)$$

$$\frac{\partial \hat{Q}}{\partial b} = x \hat{Q} \hat{P} . \qquad (296)$$

8.2 <u>The Logit transformation and the computational procedure:</u>

According to the principle of least squares, the sum of the weighted squared differences $\Sigma w_i(q_i-\hat{Q}_i)^2$ is to be minimized with respect to a and b where the weight w is taken as inversly proportional to Var q_i($= P_iQ_i/n_i$).

If instead of the observation q_i we deal with $\ell_i = \ell n \frac{p_i}{q_i}$, then for not too large differences $(q_i-Q_i)^2$ is approximately

$$(P_iQ_i)^2(\ell_i-\hat{L}_i) \tag{297}$$

(because we have $\ell n\, x = \frac{x-1}{x} + \frac{1}{2}(\frac{x-1}{x})^2$, provided $x > \frac{1}{2}$, and x here is $\frac{p_i}{q_i} / \frac{P_i}{Q_i}$).

For another approximation we might write

$$(q_i-\hat{Q}_i)^2 = (\hat{P}_i\hat{Q}_i)(p_iq_i)(\ell_i-\hat{L}_i)^2 \quad . \tag{298}$$

If now, since we do not know the true values PQ, we define the least square weights in terms of the fitted function, we have, using (298),

$$\chi^2 = \Sigma \frac{n_i}{\hat{P}_i\hat{Q}_i}(q_i-\hat{Q}_i)^2 = \Sigma\, n_i\, p_i\, q_i(\ell_i-\hat{L}_i)^2 \quad . \tag{299}$$

The weights in (299) do not contain the fitted values $\hat{P}\hat{Q}$, so that a least square solution can be obtained directly in spite of the fact that the function (292) is not linear in the parameters. Since $\hat{L}$ = a+bx the problem is reduced to that of obtaining a least square solution of a linear function with weights $w_i = n_i\, p_i\, q_i$ for ℓ_i. The method of solution then is as follows:

(a) For each observation q_i at x_i, the logit $\ell_i = \ell n \frac{1-q_i}{q_i}$ is calculated.

(b) Compute $w_i = n_i \, p_i \, q_i$ and then the quantities

$$\Sigma w,\ \Sigma wx,\ \overline{x},\ \Sigma w\ell,\ \overline{\ell},\ \Sigma wx^2,\ \frac{(\Sigma wx)^2}{\Sigma w},\ \Sigma wx\ell,\ \frac{\Sigma wx\ell \Sigma w\ell}{\Sigma w}.$$

(c) Compute:

$$a = \overline{\ell} - b\overline{x},$$

$$b = \frac{\Sigma wx\ell - \Sigma wx\Sigma w\ell/\Sigma w}{\Sigma wx^2 - (\Sigma wx)^2/\Sigma w},$$

$$\hat{L} = a - bx,$$

$$\ell n\ \hat{L.D.}\ 50 = \frac{a}{b},$$

$$\hat{Q} = \frac{1}{1 + e^{\hat{L}}},$$

$$\hat{P} = 1 - \hat{Q},$$

and

$$\chi^2 = \Sigma \frac{n_i}{\hat{P}_i \hat{Q}_i} (q_i - \hat{Q}_i)^2 .$$

9. <u>Ranked data</u>:

Another use of transformation arises in connection with ranked data.

It is often necessary to draw statistical conclusions from data giving the order of a number of magnitudes without knowing their quantitative values. Thus in psychological tests, subjects can often express preferences without being able to assign numerical values to the force with which the preference is felt. A transformation has been used to apply the analysis of variance in such cases. The observations are obtained by ranking objects given the various treatments in order of preference, say, 1,2,3...n, if there are n treatments. The procedure is to replace the rank r, say, by the expected value of the *r*th largest of a sample of size n from a normal N(0,1). These expected values have been tabulated by Fisher and Yates (1953). From the order in which any subject places a series of objects, scores may now be assigned using these tables.

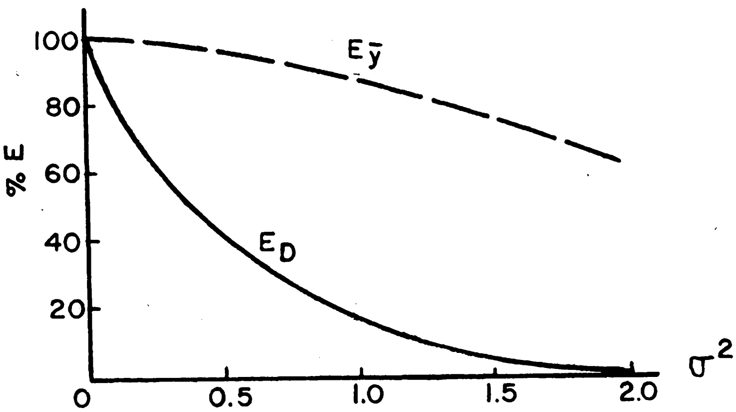

Fig. VII The Efficiency of $\bar{y}$ and D in Large Samples

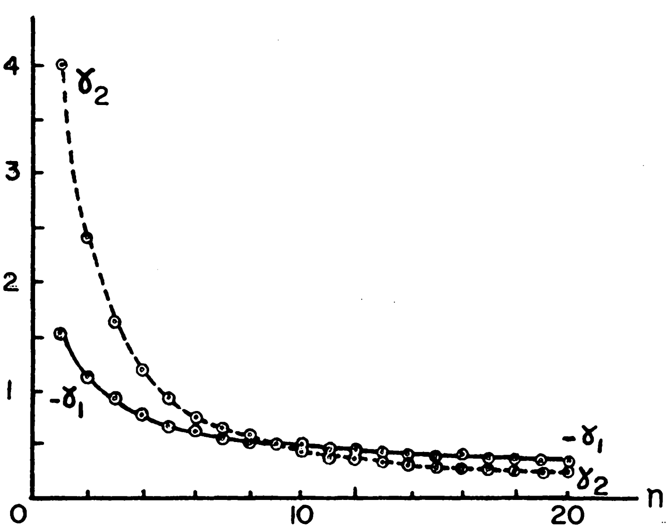

Fig. VIII Variation of γ_1, γ_2 with n

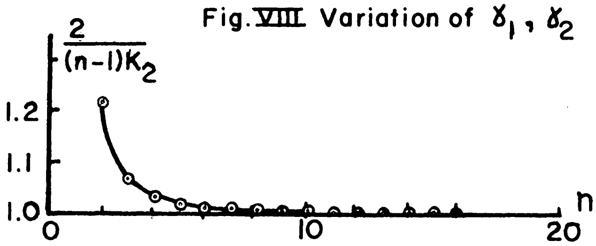

Fig. IX Variation of $\frac{2}{(n-1)K_2}$ with n

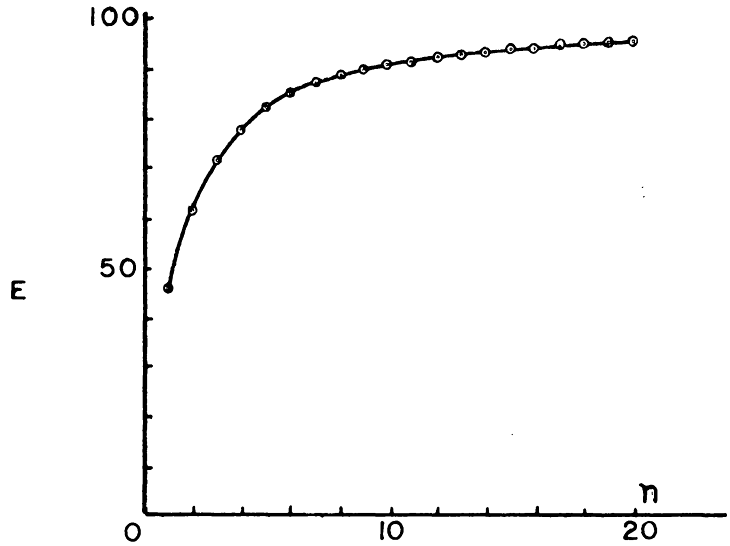

Fig. X The Efficiency of the Mean of Ln S^2 for Estimating σ^2

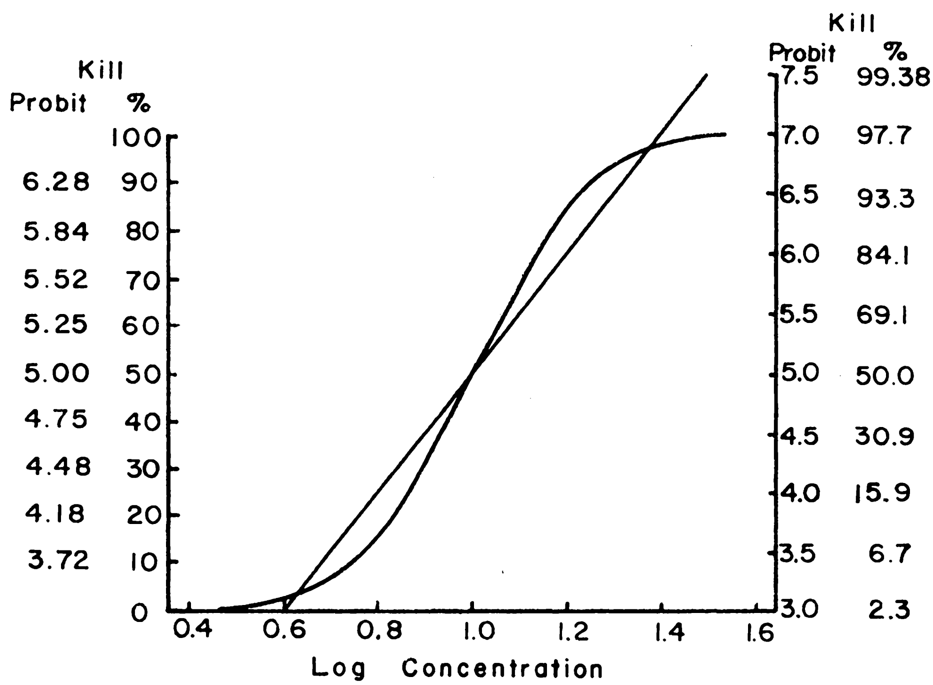

Fig. XI Effect of the Probit Transformation

10. Acknowledgement.

I wish to express my sincere appreciation and my recognition to Professor B. Harshbarger for his helpful encouragement and to Professor L. R. Shenton for his valuable guidance during the course of this work.

11. References.

Anscombe, F. J. (1948). The transformation of Poisson, Binomial, and Negative Binomial data. Biometrika 35, 246.

Bartlett, M. S. (1936). Square root transformation in analysis of variance. J.R.S.S. Suppl. 3, 68.

_____, (1947). The use of transformations. Biometrics 3, 39.

_____, and Kendall, D. G. (1946). The statistical analysis of variance heterogeneity and the logarithmic transformation. J.R.S.S. Suppl. 8, 128.

Beall, G. (1942). Transformation of data from entomological field experiment so that the ANOVA is applicable. Biometrika 32, 243.

Berkson, J. (1944). Application of the logistic function to bioassay. J.A.S.A. 39, 357.

Bliss, C. I. (1935). The calculation of dosage-mortality curve. Annals of Applied Biology 22, 134.

_____, (1938). The transformation of percentages for use in ANOVA. Ohio Journal of Science, 38, 9.

Blom, G. (1954). Transformation of the Binomial, Negative Binomial, Poisson, and χ^2 distributions. Biometrika, 41.

Curtiss, J. H. (1943). The use of transformations in ANOVA. A.M.S. 14, 107.

Cramer, H. (1946). Mathematical methods of statistics. Princeton University Press.

_____, (1937). Random variables and probability distributions. Cambridge.

Cornish and Fisher (1937). Contributions to Mathematical Statistics. New York, Wiley and Sons.

Cochran, W. G. (1940). The ANOVA when experimental errors follow the Poisson or Binomial distribution. A.M.S. 11, 335.

_____, (1947). Some consequences when the assumptions of ANOVA are not satisfied. _Biometrics_ 3.

Clark, Andrew and others (1939). The ANOVA with special reference to data expressed in percentages. _J. of the American Society of Agronomy_ 31, 55.

David, F. N. (1955). The transformation of discrete variables. _Annals of Human Genetics_ 19, 174.

_____, (1938). _Tables of the correlation coefficient_. Biometrika Office, University College, London.

Eisenhart, C., and others (1947). _Selected techniques of statistical analysis_. McGraw Hill, New York.

Eisenhart, C. (1947). The assumptions underlying the ANOVA. _Biometrics_ 3.

Evans, G. C. (1927). _The logarithmic potential_. New York.

Fisher, R. A. (1950). _Statistical methods for research workers_. 11th edd.

_____, (1949). _The design of experiments_. 5th edd.

_____, and Yates (1953). _Statistical Tables_.

Freeman, M. F. and Tukey, J. W. (1950). Transformations related to the angular and the square root. A.M.S. 21, 607.

Finney, D. J. (1941). On the distribution of a variate whose logarithm is normally distributed. J.R.S.S. Suppl. 7, 155.

_____, (1947). _Probit analysis_. Cambridge University Press.

Kempthorn, O. (1951). _Design of experiment_. New York, Wiley and Sons. 153.

Kendall, M. G. (1952). Advanced theory of statistics, Vol. I. New York, Hafner Co.

_____, (1951). Advanced theory of Statistics, Vol. II. New York, Hafner Co.

Molina, E. C. (1947). Poisson's exponential Binomial limits. New York, Van Nostrand Co.

Peiser, A. M. (1943). A.M.S. 20, 417.

Pearson, E. S. and Hartley, H. O. (1958). Biometrika tables for statisticians. Cambridge University Press.

Scheffe, H. (1958). The analysis of variance. New York, Wiley and Sons.

Stevens, W. L. (1950). Fiducial limits of the parameter of a discontinuous distribution. Biometrika 37, 117.

Tukey, J. W. (1949). One degree of freedom for non-additivity. Biometrics 5.

Thompson, C. M. (1941). Incomplete Beta-function. Biometrika 32, 151.

Taylor, R. J. (1955). On the use of an auxiliary variable in the transformation of discrete data. M.S. thesis, V.P.I.

Uspensky, J. V. (1937). Introduction to mathematical probability. New York.

Williams, C. B. The use of logarithm in the interpretation of certain entomological problems. Annals of Applied Biology 24, 404.

Wilson, E. B. and Hilferty, M. M. (1931). The distribution of Chi-square. Proc. Nat. Acad. Sci. 17, 694.

Winsor, C. P. (1932). A comparison of certain symmetrical growth curves. J. of the Washington Academy of Sciences 22, 73.

Widder, D. V. (1961). Advanced Calculus. Prentice-Hall, Inc.

Yule, C. U. and Kendall, M. G. (1950). An Introduction to the Theory of Statistics. Hafner.

Zubin, J. (1936). Note on a transformation function for proportions and percentages. Journal of Applied Psychology 19, 213.

Further References

Anscombe, F. J. (1949). The statistical analysis of insect counts based on the Negative Binomial distribution. Biometrics 5, 165.

Bartlett, M. S. (1937). Some examples of statistical methods of research in agriculture and applied biology. J.R.S.S. Suppl. 4.

Bliss, C. I. (1935). The comparison of dosage mortality data. Annals of Applied Biology 22, 307.

Cohen, A. C. (1951). Estimating parameters of logarithmic normal distributions by maximum likelihood. J.A.S.A. 46, 206.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. J.R.S.S. Suppl. 4, 102.

Cochran, W. G. and Carroll, S. P. (1953). A sampling investigation of the efficiency of weighting inversely as the estimated variance. Biometrics 9, 447.

Cochran, W. G. (1938). Some difficulties in the statistical analysis of replicated experiments. Empire J. of Experimental Agriculture 6, 157.

______, (1943). ANOVA of percentages based on unequal numbers. J.A.S.A. 38, 287.

Ghurys, G. (1949). Transformation of a Binomial variate for the ANOVA. _J. Indian Soc. Agric. Stat._ 2, 94.

Reed, L. J. and Berkson, J. (1929). The application of the Logistic function to experimental data. _J. of Physical_ Chemistry 33, 760.

The vita has been removed from the scanned document

ABSTRACT

This paper is a review of the major literature dealing with transformations of random variates which achieve variance stabilization and approximate normalization. The subject can be said to have been initiated by a genetical paper of R. A. Fisher (1922) which uses the angular transformation $\phi = 2 \arcsin\sqrt{p}$ to deal with the analysis of proportions p with E(p) = P. Here it turns out that Var ϕ is almost independent of P and so stabilizes the variance. Some fourteen years later Bartlett introduced the so-called square-root transformation which achieves variance stabilization for variates following a Poisson distribution. These two transformations and their ramifications in theory and application are fully discussed along with refinements introduced by later writers, notably Curtiss (1943) and Anscombe (1948).

Another important transformation discussed is one which refers an analysis of observations on to a logarithmic scale, and here there are uses in analysis of variance situations and theoretical problems in the field of estimation; in the case of the latter, the work of D. J. Finney (1941) is considered in some detail. The asymptotic normality of the transformation is also considered.

Transformations primarily designed to bring about ultimate normality in distribution are also included. In particular there is reference to work on the chi-square probability integral (Fisher), (Wilson and Hilferty (1931)) and the logarithmic transformation of a correlation coefficient (Fisher (1921)).

Other miscellaneous topics referred include

i. the probability integral transformation (Probits), with applications in bioassay;

ii. applications of transformation theory to set up approximate confidence intervals for distribution parameters (Blom (1954));

iii. transformations in connection with the interpretation of so-called 'ranked' data.