1954

An extension of preliminary tests of significance permitting control of disturbances in statistical inferences

David Vernon Huntsberger

Iowa State College

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation

Huntsberger, David Vernon, "An extension of preliminary tests of significance permitting control of disturbances in statistical inferences" (1954). Retrospective Theses and Dissertations. 12863.
https://lib.dr.iastate.edu/rtd/12863

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
NOTE TO USERS

This reproduction is the best copy available.

UMI®
AN EXTENSION OF PRELIMINARY TESTS OF SIGNIFICANCE PERMITTING CONTROL OF DISTURBANCES IN STATISTICAL INFERENCES

by

David Vernon Huntsberger

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of The Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State College

1954
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
TABLE OF CONTENTS

I. INTRODUCTION 1

II. A GENERALIZED POOLING PROCEDURE 8
   A. The Problem 8
   B. Rule of Procedure 8
   C. A Criterion 9

III. POOLING SAMPLE MEANS, A SPECIAL CASE 10
   A. The Problem 10
      1. Statement of the problem 10
      2. Rule of procedure 10
   B. Derivation of the Mean Square Deviation and the Bias 11
      1. The mean square deviation of \(W(t)\) 11
      2. The bias of \(W(t)\) 12
   C. A Weighting Function, \(\phi_o(t)\) 13
   D. Evaluation of the Mean Square Deviation 14
      1. \(E \left[ W(t) - \mu_1 \right]^2 \) when \(\phi(t) = \phi_o(t)\) 14
      2. \(E \left[ W(t) - \mu_1 \right]^2 \) for the "sometimes pool" procedure, \(\text{SP}(t)\) 23
   E. Evaluation of the Bias 24
      1. The bias of \(W_o(t)\) 24
      2. The bias of \(\text{SP}(t)\) 25
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>F. Results</td>
<td>26</td>
</tr>
<tr>
<td>1. Tables and graphs</td>
<td>26</td>
</tr>
<tr>
<td>2. Comparison of the efficiencies of $W_0(t)$ and $SP(t)$</td>
<td>36</td>
</tr>
<tr>
<td>3. Effect of sample size and variance on the efficiency of $W_0(t)$</td>
<td>36</td>
</tr>
<tr>
<td>4. Comparison of the biases of $W_0(t)$ and $SP(t)$</td>
<td>37</td>
</tr>
<tr>
<td>IV. OTHER CASES OF POOLING MEANS</td>
<td>38</td>
</tr>
<tr>
<td>A. Different Known Variances</td>
<td>38</td>
</tr>
<tr>
<td>1. Statement of the problem</td>
<td>38</td>
</tr>
<tr>
<td>2. Rule of procedure</td>
<td>38</td>
</tr>
<tr>
<td>3. Derivation of the mean square deviation of $W(t)$</td>
<td>39</td>
</tr>
<tr>
<td>B. Some Special Cases</td>
<td>42</td>
</tr>
<tr>
<td>1. Equal population variances, different sample sizes</td>
<td>42</td>
</tr>
<tr>
<td>2. Different population variances, equal sample sizes</td>
<td>42</td>
</tr>
<tr>
<td>3. Different population variances, different sample sizes</td>
<td>43</td>
</tr>
<tr>
<td>C. Results</td>
<td>44</td>
</tr>
<tr>
<td>V. POOLING NORMAL OR NEAR NORMAL ESTIMATORS</td>
<td>46</td>
</tr>
<tr>
<td>A. Normally Distributed Estimators</td>
<td>46</td>
</tr>
<tr>
<td>B. Approximately Normal Estimators</td>
<td>48</td>
</tr>
<tr>
<td>VI. WEIGHTING FUNCTIONS</td>
<td>50</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----</td>
</tr>
<tr>
<td>VII. A TWO-PARAMETER FAMILY OF WEIGHTING FUNCTIONS</td>
<td>57</td>
</tr>
<tr>
<td>A. The Family Defined</td>
<td>57</td>
</tr>
<tr>
<td>B. The Mean Square Deviation of $W_{ab}(t)$</td>
<td>57</td>
</tr>
<tr>
<td>C. Results</td>
<td>61</td>
</tr>
<tr>
<td>1. Tables</td>
<td>61</td>
</tr>
<tr>
<td>2. Effects of varying $b$ when $a$ is held constant</td>
<td>61</td>
</tr>
<tr>
<td>3. Effects of varying $a$ when $b$ is held constant</td>
<td>65</td>
</tr>
<tr>
<td>4. Comparison of the efficiencies of $W_o(t)$, $W_{ab}(t)$, and $SP(t)$</td>
<td>65</td>
</tr>
<tr>
<td>VIII. TRANSFORMED CORRELATION COEFFICIENTS</td>
<td>69</td>
</tr>
<tr>
<td>A. The Transformation</td>
<td>69</td>
</tr>
<tr>
<td>B. The Pooling Procedure</td>
<td>70</td>
</tr>
<tr>
<td>C. Examples</td>
<td>72</td>
</tr>
<tr>
<td>IX. DISCUSSION</td>
<td>76</td>
</tr>
<tr>
<td>X. SUMMARY</td>
<td>80</td>
</tr>
<tr>
<td>XI. LITERATURE CITED</td>
<td>83</td>
</tr>
<tr>
<td>XII. ACKNOWLEDGEMENTS</td>
<td>85</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

In the statistical analysis of quantitative data the observed values, $X_1, X_2, \ldots, X_n$, are assumed to have a joint probability density function

$$f(X_1, \ldots, X_n; \theta_1, \ldots, \theta_k),$$

(1)

where the functional form, $f$, is known and the $\theta$'s are parameters which may or may not be known.

Statistical theory provides procedures for estimating the parameters and for testing the hypothesis that the data follow a more specialized probability density function

$$f_1(X_1, \ldots, X_n; \theta'_1, \ldots, \theta'_k),$$

(2)

where $f_1$ is some member or subgroup of the family $f$. These procedures give estimates and tests which are "best" in some sense.

It often happens, however, that there is reason to doubt the appropriateness of the specifications in (1) and (2). When such doubt arises it is not uncommon that an attempt is made to resolve the uncertainty by considering the questionable assumption as a hypothesis to be tested. Depending upon the outcome of the test, the specifications remain unchanged or are altered to conform with the evidence obtained from the test, and techniques consistent with the decision are used to perform the final analysis.

If the test of a specification and the final analysis are performed on separate sets of data, the statistical inferences are perfectly valid;
but they must be prefaced by a qualifying phrase, such as "If the assumptions are correct, then ...." If, on the other hand, the test of the specification and the final analysis are performed on the same set of data, as is common, the situation becomes complex as the final inferences are not statistically independent of the test. In this latter case the test of the specification is referred to as a preliminary test of significance or preliminary test.

Since a preliminary test is a statistical test of a hypothesis, it is subject to errors of the first and second kinds, and these errors introduce disturbances into the statistical inferences based on the final analysis of the data. The disturbances may result in biased estimates and a loss of efficiency as regards estimation, or they may cause shifts in the size and power of a test of a hypothesis. Since the motivation for the use of a preliminary test is often the intuitive feeling that an increase in efficiency or power may result, it is advisable to determine whether or not such a gain may, in fact, be achieved. Despite the fact that the effects of preliminary testing have been studied in only a few special cases, many examples of the use of such tests in a wide variety of applications may be found in the literature of Statistics.

In the first detailed study of the effects of preliminary tests, Bancroft (1) investigated the bias, variance, and mean square error of an estimate of a variance based on a preliminary test of the equality of two variances. The estimator employed was determined by the following rule of procedure.

Given two independent estimates, \( S_1^2 \) and \( S_2^2 \), for \( \sigma_1^2 \) and \( \sigma_2^2 \) respectively, test the hypothesis \( H_0: \sigma_1^2 = \sigma_2^2 \) by the F-test where the critical region is defined as
\[
\frac{S_1^2}{S_2^2} > F_{n_2}^{n_1} (a),
\]

and \( F_{n_2}^{n_1} (a) \) is the a\% point of the F distribution with \( n_1 \) and \( n_2 \) degrees of freedom. If the calculated ratio falls into the critical region, use \( S_1^2 \) to estimate \( \sigma_1^2 \). If the ratio falls into the acceptance region, use

\[
\frac{n_1S_1^2 + n_2S_2^2}{n_1 + n_2}
\]

to estimate \( \sigma_1^2 \).

In this investigation it was found that the bias and mean square error depend upon the value of the ratio of the two variances,

\[
\frac{\sigma_2^2}{\sigma_1^2},
\]

and that in the cases for which these quantities were evaluated there is no value of \( a \) which minimizes the bias for all values of

\[
\frac{\sigma_2^2}{\sigma_1^2}.
\]

Further, it was found that if \( a \) is chosen so that \( F_{n_2}^{n_1} (a) = 1 \), the mean square error is smaller than the variance of \( S_1^2 \) except for small values of

\[
\frac{\sigma_2^2}{\sigma_1^2};
\]

but the mean square error is always larger than the mean square error of

\[
\frac{n_1S_1^2 + n_2S_2^2}{n_1 + n_2}
\]

except for \( \frac{\sigma_2^2}{\sigma_1^2} = 1 \).
The same paper considered the problem of the bias introduced into the estimate, $b_1$, for a regression coefficient, $\beta_1$, in the regression equation, $Y = b_0 + b_1X_1 + b_2X_2$, when the decision to retain or delete the independent variable $X_2$ is based on a test of the significance of $b_2$. An expression for the bias was obtained.

Bancroft (2) also reported on biases in estimates of variance in ordinary multiple regression due to the deletion of independent variables.

Mosteller (10) considered a similar problem concerning the pooling of sample means where the samples are assumed to be independent random samples from similar normal populations with common, known variance, and a preliminary test of the equality of the population means determines whether the sample mean, $\bar{X}_1$, or the pooled estimate,

$$\frac{\bar{X}_1 + \bar{X}_2}{2}$$

is used to estimate $\mu_1$, the mean of the first population. In order to compare the efficiency of the estimate resulting from the preliminary test procedure with that of the sample mean, $\bar{X}_1$, he computed a disadvantage coefficient, $C$, the ratio of the mean square deviation of the sometimes pool estimator to the variance of $\bar{X}_1$. He showed that $C$ is a function of the nuisance parameter, $\gamma = (\mu_1 - \mu_2) \sqrt{\frac{n}{2}} \sigma$, and that regardless of the significance level of the preliminary test there is a range of values of $\gamma$ for which the "sometimes pool" procedure can lead to a considerable loss of efficiency in cases where a gain is anticipated.
In the analysis of variance there may be uncertainty as to whether two mean squares are homogeneous. If they do, in fact, estimate the same variance, it is desirable to pool them in order to obtain the best estimate of variance. The common practice is to test for homogeneity of the variance estimates and if the test is non-significant at a specified level to use the pooled estimate as the denominator of the F-test of the major hypothesis. In the event that this preliminary test is significant, the appropriate mean square, as indicated by the expected mean squares in the analysis, is used as the denominator without recourse to pooling.

The effects of this type of preliminary test on the size and power of the analysis of variance test of the major hypothesis have been studied independently by Paull (11) and Bechhofer (3). Paull considered a components of variance model, while Bechhofer was concerned with a more general class of problems; those which can be represented as a test of a general linear hypothesis in canonical form. Both of these investigations indicate that when no information regarding the nuisance parameters is available there is little justification for using the preliminary test. The advantages that might be gained if the nuisance parameters were small are outweighed by the disturbances that could result if these parameters are intermediate in value.

Kitagawa (9) derived the distribution function and the moments for the estimator obtained from the rule of procedure studied by Bancroft in the variance estimation problem; and also derived the distribution and moments for a pooled estimate of a mean based on a preliminary
test when the variance is unknown.

The work of Mosteller and Kitagawa was extended by Bennett (4) to cases where preliminary tests are performed for both homogeneity of variances and equality of means prior to estimating the mean or testing hypotheses about the mean. He derived the distribution functions, the biases, and the mean square errors for different situations depending upon what could be assumed known regarding the parameters of the associated normal distributions. Included in this study was the power function for the t-test of the equality of means subsequent to an F-test for homogeneity of variance.

All preliminary tests, the special cases which have been investigated and the many others which are common practice, have certain characteristics in common. A statistic, $T$, is evaluated using the data at hand. If $T$ is not significant at the preassigned significance level $\alpha$, that is, if $T$ falls into the region of acceptance, a pooled statistic, $K_1$, is used to estimate the parameter or is employed in testing a hypothesis about the parameter. If $T$ is significant, that is, if $T$ falls into the region of rejection, the "never pool" statistic, $K_2$, is used as the estimator or is used in testing the hypothesis. In any event, the only information derived from $T$ is that it does or does not fall into the region of rejection. It would appear that if more of the information contained in the value of $T$ were to be utilized, it should be possible to exert more control over the disturbances inherent in preliminary testing than is possible by merely shifting the level of significance of the preliminary test.
It is the purpose of this investigation to formulate a generalized pooling procedure, which includes the "sometimes pool" procedure based on the preliminary test as a special case, and to investigate the operating characteristics of this procedure when it is utilized in a particular class of problems. The first problem selected for study is the one considered by Mosteller, that of pooling sample means. This particular problem was selected because it is one of the least complex; it was felt that the implications would be more easily understood here than in a more complex situation, and that considerable insight into the pooling problem could be gained by examination of an elementary problem. The results obtained are generalized to include a larger class of estimation problems.
II. A GENERALIZED POOLING PROCEDURE

A. The Problem

Let \( X_1, \ldots, X_n \) be a random sample with joint probability density function

\[
f(X_1, \ldots, X_n; \theta_1, \theta_2, \ldots, \theta_k),
\]

where the functional form is known, \( \theta_1 \) and \( \theta_2 \) are unknown parameters, and the last \( k-2 \) \( \theta \)'s are parameters whose values may or may not be known. Let \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) be the usual estimators for \( \theta_1 \) and \( \theta_2 \) as provided by statistical theory. These estimators are best in some sense.

If we know that \( \theta_1 = \theta_2 \) we can pool the information about \( \theta_1 \) contained in both \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) and use a function, \( g(\hat{\theta}_1, \hat{\theta}_2) \), to estimate \( \theta_1 \). Such a procedure will, in general, yield estimators which are better than \( \hat{\theta}_1 \) according to the same or different criteria of goodness.

When we do not know whether or not \( \theta_1 = \theta_2 \), we may still feel that a better estimator for \( \theta_1 \) may be found by utilizing any information \( \hat{\theta}_2 \) may provide concerning the value of \( \theta_1 \).

B. Rule of Procedure

Let \( T \) be the statistic which the theory indicates will provide the best test of the hypothesis \( H_0: \theta_1 = \theta_2 \) against the alternative hypothesis \( H_a: \theta_1 \neq \theta_2 \).
Using the sample data, evaluate $T$.

To estimate $\theta_1$ use the function

$$W(T) = \phi(T) \hat{\theta}_1 + 1/T - \phi(T) \gamma \frac{g(\hat{\theta}_1, \hat{\theta}_2)}{g(\hat{\theta}_1, \hat{\theta}_2)},$$

where $\phi(T)$ is a function of $T$ only, and $g(\hat{\theta}_1, \hat{\theta}_2)$ is the pooled estimator that is used when it is known that $\theta_1 = \theta_2$.

If $\phi(T)$ is defined as

$$\phi(T) = 0 \quad , \quad T \leq A_a,$$

$$\phi(T) = 1 \quad , \quad T \leq R_a,$$

where $A_a$ and $R_a$ are the acceptance and rejection regions for the test of $H_0: \theta_1 = \theta_2$ with probability of type I error equal to $a$, then $W(T)$ becomes the estimator based on the ordinary preliminary test of significance.

Thus the preliminary test procedure is a special case of the rule of procedure outlined above.

C. A Criterion

In order to determine whether or not the estimator $W(T)$ has any advantages over the estimator $\hat{\theta}_1$, it is necessary to select a criterion on which we can base our judgments. It is readily apparent that $W(T)$ will, in general, provide biased estimates of $\theta_1$ while in many cases $\hat{\theta}_1$ will give unbiased estimates. If we insist on unbiasedness, the problem is solved for we will always use $\hat{\theta}_1$ unless it is known that $\theta_1 = \theta_2$. The criterion which we shall use is the mean square deviation, $D^2$, from the true parameter value, $\theta_1$, the smaller value being associated with the better estimator.
III. POOLING SAMPLE MEANS, A SPECIAL CASE

A. The Problem

1. Statement of the problem

Given two independent random samples of size $n$, one from each of two normal populations which have the same known variance, $\sigma^2$, and means $\mu_1$ and $\mu_2$ respectively, an estimate of $\mu_1$ is obtained from the sample means, $\bar{X}_1$ and $\bar{X}_2$, by following the rule of procedure stated in Section II. The mean square deviation of this estimator is compared with that of the estimator based on the preliminary test of the hypothesis of equal population means. The biases of these estimators are also compared.

2. Rule of procedure

For testing the hypothesis that $\mu_1 = \mu_2$ the appropriate statistic is the standard normal deviate

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sigma} \sqrt{\frac{n}{2}}$$  \hspace{1cm} (1)

which has expectation

$$\chi = \mathbb{E}(t) = \frac{\mu_1 - \mu_2}{\sigma} \sqrt{\frac{n}{2}}$$  \hspace{1cm} (2)

To estimate $\mu_1$, use

$$W(t) = \phi(t) \bar{X}_1 + (1 - \rho(t)) \frac{\bar{X}_1 + \bar{X}_2}{2}$$  \hspace{1cm} (3)
where \( \varphi(t) \) is a function of \( t \) only.

If we define \( \varphi(t) \) as

\[
\varphi(t) = 0 \quad \text{when} \quad |t| \leq t_a,
\]
\[
\varphi(t) = 1 \quad \text{when} \quad |t| > t_a,
\]

where \( P(|t| > t_a) = \alpha \), then \( W(t) \) reduces to the estimator based on the preliminary test.

B. Derivation of the Mean Square Deviation
and the Bias

1. The mean square deviation of \( W(t) \)

The estimator, \( W(t) \) defined in (3) may also be written as

\[
W(t) = \frac{\bar{X}_1 + \bar{X}_2}{2} + \varphi(t) \frac{\bar{X}_1 - \bar{X}_2}{2},
\]

or, if \( a = \sqrt{\frac{2}{n}} \), as

\[
W(t) = \frac{\bar{X}_1 + \bar{X}_2}{2} + \frac{at \varphi(t)}{2}.
\]

Since \( \varphi(t) \) is a function of \( t \) only and \( t \) and \( \bar{X}_1 + \bar{X}_2 \) are independently distributed, the mean square deviation of \( W(t) \) is

\[
D_W^2(y') = E \left[ (W(t) - \mu_1)^2 \right] = \frac{1}{4} E \left[ (\bar{X}_1 + \bar{X}_2)^2 \right] + \frac{a^2}{4} E \left[ t^2 \varphi^2(t) \right] + \frac{a}{2} E \left[ \bar{X}_1 + \bar{X}_2 \right] E \left[ t \varphi(t) \right] - \mu_1 E \left[ \bar{X}_1 + \bar{X}_2 \right] - \mu_1 E \left[ t \varphi(t) \right] + \mu_1^2.
\]

Substituting for the expectations in the right hand member,
\[ D^2_W(\gamma) = \frac{\sigma^2}{2n} + \left( \frac{\mu_1 - \mu_2}{2} \right)^2 + \frac{\sigma^2}{2n} \int_{-\infty}^{\infty} t^2 \phi^2(t) N(t - \gamma) dt \]

\[ - \frac{\sigma}{\sqrt{n}} \frac{\mu_1 - \mu_2}{2} \int_{-\infty}^{\infty} t(\theta)(t) N(t - \gamma) dt, \]

where \( N(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \).

Further simplification yields

\[ D^2_W(\gamma) = \frac{\sigma^2}{2n} \left\{ 1 + \int_{-\infty}^{\infty} \left[ t\theta(t) - \gamma \right]^2 N(t - \gamma) dt \right\}. \quad (4) \]

2. The bias of \( W(t) \)

The expectation of \( W(t) \) is

\[ E\left[ W(t) \right] = \frac{1}{2} E\left[ X_1 + X_2 \right] + \frac{a}{2} \int_{-\infty}^{\infty} t\theta(t) N(t - \gamma) dt, \]

\[ = \frac{\mu_1 + \mu_2}{2} + \frac{a}{2} \int_{-\infty}^{\infty} t\theta(t) N(t - \gamma) dt, \]

so that the bias, \( \mu_1 - E\left[ \sqrt{W(t)} \right] \) is

\[ B_W(\gamma) = \frac{\mu_1 - \mu_2}{2} - \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} t\theta(t) N(t - \gamma) dt. \quad (5) \]
C. A Weighting Function, \( \varphi_o(t) \)

For a given value of \( \gamma \) the mean square deviation of \( W(\gamma) \) can be expressed as

\[
D^2 = \frac{\sigma^2}{2n} \left[ 1 + \varphi^2 + (1-\varphi)^2 \gamma^2 \right],
\]

where \( \varphi \) is a function of \( \gamma \). If \( D^2 \) is minimized with respect to \( \varphi \)

\[
\frac{\partial D^2}{\partial \varphi} = \frac{\sigma^2}{2n} \left[ \varphi - (1-\varphi) \right]^2 = 0,
\]

and

\[
\varphi' = \frac{\gamma^2}{1 + \gamma^2}
\]

\( \varphi' \) is that function of \( \gamma \) which, when substituted for \( \varphi \) in \( D^2 \), provides the minimum value of \( D^2 \). This is seen to be a minimum since

\[
\frac{\partial^2 D^2}{\partial \varphi^2} \bigg|_{\varphi=\varphi'} = \frac{\sigma^2}{2n} (1 + \gamma^2) > 0.
\]

Since \( \gamma \) is generally unknown and since \( t \) is an unbiased estimator for \( \gamma \), it was decided to estimate \( \varphi' \) by

\[
\varphi_o(t) = \frac{t^2}{1+t^2}, \quad (6)
\]
D. Evaluation of the Mean Square Deviation

1. \( E \left[ W(t) - \mu_1 \right]^2 \) when \( \phi(t) = \phi_o(t) \)

If \( \phi_o(t) \) is substituted for \( \phi(t) \), the mean square deviation of \( W(t) \) becomes

\[
D_{\phi_0}^2 (\gamma) = \frac{\sigma^2}{2n} \left\{ 1 + \int_{-\infty}^{\infty} \left[ \gamma - \frac{t^3}{1+t^2} \right]^2 N(t-\gamma) dt \right\}.
\]

Making use of the fact that

\[
\frac{t^3}{1+t^2} = t - \frac{t}{1+t^2}
\]

the integral can be written as

\[
I(\gamma) = \int_{-\infty}^{\infty} \left[ (t - \gamma) - \frac{t}{1+t^2} \right]^2 N(t-\gamma) dt,
\]

\[
I(\gamma) = \int_{-\infty}^{\infty} \left[ (t - \gamma)^2 - \frac{2t(t - \gamma)}{1+t^2} + \frac{t^2}{(1+t^2)^2} \right] N(t-\gamma) dt,
\]

\[
I(\gamma) = 1 - 2 \int_{-\infty}^{\infty} \frac{t(t - \gamma)}{1+t^2} N(t-\gamma) dt + \int_{-\infty}^{\infty} \frac{t^2}{(1+t^2)^2} N(t-\gamma) dt.
\]

(7)
The first integral of (7) may be integrated by parts.

Let:

\[ u = \frac{t}{1 + t^2}, \quad du = \frac{1 - t^2}{(1 + t^2)^2} dt, \]

\[ v = -N(t - y), \quad dv = (t - y)N(t - y)dt. \]

then

\[ \int_{-\infty}^{\infty} \frac{t(t - y)}{1 + t^2} N(t - y) dt = \left[ -\frac{t}{1 + t^2} N(t - y) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1 - t^2}{(1 + t^2)^2} N(t - y), dt, \]

\[ = \int_{-\infty}^{\infty} \frac{1 - t^2}{(1 + t^2)^2} N(t - y) dt. \]

Substituting into (7),

\[ I(y) = 1 - 2 \int_{-\infty}^{\infty} \frac{1}{(1 + t^2)^2} N(t - y) dt + 3 \int_{-\infty}^{\infty} \frac{t^2}{(1 + t^2)^2} N(t - y) dt, \]

or

\[ I(y) = 1 + 3 \int_{-\infty}^{\infty} \frac{1}{1 + t^2} N(t - y) dt - 5 \int_{-\infty}^{\infty} \frac{1}{(1 + t^2)^2} N(t - y) dt. \]

This can be expressed symbolically as

\[ I(y) = 1 + 3 H(y) - 5 G(y). \quad (8) \]

The integral \( H(y) \) was evaluated using a method given in a paper by Zemansky (13).
\[ H(\chi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} (t-\chi)^2} \frac{dt}{1 + t^2} \]

\[ = \frac{e^{-\chi^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \frac{\cos i\chi t}{1 + t^2} \, dt, \]

\[ H(\chi) = \sqrt{\frac{2}{\pi}} e^{-\chi^2/2} \int_{0}^{\infty} e^{-u} \left(1 + u\right) \cos i\chi t \, du, \]

Since

\[ \frac{1}{1 + t^2} = \int_{0}^{\infty} e^{-(1+t^2)u} \, du, \]

\[ H(\chi) = \sqrt{\frac{2}{\pi}} e^{-\chi^2/2} \int_{0}^{\infty} e^{-u} \left(1 + u\right) \cos i\chi t \, du, \]

or

\[ H(\chi) = \frac{e^{-\chi^2/2}}{\sqrt{2}} \int_{0}^{\infty} e^{-u + \frac{\chi^2}{2(1+2u)}} \frac{du}{\sqrt{1 + u}}. \]

Let \( u + \frac{1}{2} = y^2 \), \( du = 2y \, dy \), then
\begin{align*}
H(\gamma) &= \sqrt{2} \int_{-\frac{1}{2}(\gamma^2 - 1)}^{\infty} e^{-\gamma^2 - \frac{\gamma^2}{4y^2}} \, dy, \\
&= \frac{1}{\sqrt{2}} \\
H(\gamma) &= \sqrt{2} \int_{-\frac{\gamma^2 + 1}{2}}^{\infty} e^{-\gamma^2} \sum_{n=0}^{\infty} \frac{\gamma^{2n}}{2^n n! y^{2n}} \, dy.
\end{align*}

A transformation,

\[ z = \frac{\gamma}{\sqrt{2}}, \quad dz = \frac{dy}{\sqrt{2}}, \]

transforms \( H(\gamma) \) into,

\begin{align*}
H(\gamma) &= e^{-\gamma^2/2} \sum_{n=0}^{\infty} \frac{e^{-\gamma^2/2}}{n!} \left( \frac{\gamma^2}{2} \right)^n A_n, \\
&= \sum_{n=0}^{\infty} e^{-\gamma^2/2} \left( \frac{\gamma^2}{2} \right)^n A_n, \\
\end{align*}

where

\[ A_n = \frac{1}{\sqrt{2}} \int_{1}^{\infty} e^{-\frac{z^2}{2}} \frac{1}{z^{2n}} \, dz, \]

\[ A_0 = \frac{1}{\sqrt{2}} \int_{1}^{\infty} e^{-\frac{z^2}{2}} \, dz. \]
The evaluation of the integrals $A_n$, $n = 1, 2, 3, \ldots$, may be accomplished by means of a recursion formula.

Let

$$u = e^{-\frac{z^2}{2}}, \quad du = -ze^{-\frac{z^2}{2}} dz,$$

$$v = -\frac{1}{(2n-1)z^{2n-1}} \quad dv = \frac{1}{z^{2n}} dz,$$

then

$$A_n = e^\frac{1}{2} \left\{ \int_1^\infty \frac{-e^{-\frac{z^2}{2}}}{(2n-1)z^{2n-1}} dz - \int_1^\infty \frac{e^{-\frac{z^2}{2}}}{z^{2(n-1)}} dz \right\},$$

$$= e^\frac{1}{2} \left[ \frac{1}{2n-1} - \frac{1}{2n-1} \int_1^\infty \frac{e^{-\frac{z^2}{2}}}{z^{2(n-1)}} dz \right],$$

$$A_n = \frac{1}{2n-1} \left[ 1 - A_{n-1} \right].$$

The integral $A_o$ was evaluated to eight significant figures using tables of the cumulative normal distribution function (6).

$$A_o = \sqrt{2 \pi} \int_1^\infty e^\frac{1}{2} \left\{ \int_1^\infty N(y)dy = .65567951. \right\}$$

To evaluate the second integral in the right hand member of equation (8)

$$G(\gamma) = \int_{-\infty}^{\infty} \frac{1}{(1+t^2)^{\frac{3}{2}}} N(t-\gamma)dt,$$
If we first show that

\[ 2G(\gamma) = J'(\gamma) + H(\gamma), \]

where

\[ J'(\gamma) = \frac{dJ(\gamma)}{d\gamma} = \frac{d}{d\gamma} \left[ -\int_{-\infty}^{\infty} \frac{t}{1+t^2} N(t-\gamma) dt \right], \]

\[ J'(\gamma) = \int_{-\infty}^{\infty} \frac{t(t-\gamma)}{1+t^2} N(t-\gamma) dt. \]

\( J'(\gamma) \) may be integrated by parts as follows:

Let

\[ u = \frac{t}{1+t^2}, \quad du = \frac{1-t^2}{(1+t^2)^2} dt, \]
\[ v = -N(t-\gamma), \quad dv = (t-\gamma)N(t-\gamma) dt, \]

then

\[ J'(\gamma) = \left[ -\frac{t}{1+t^2} N(t-\gamma) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1-t^2}{(1+t^2)^2} N(t-\gamma) dt, \]

or

\[ J'(\gamma) = \int_{-\infty}^{\infty} \frac{1-t^2}{(1+t^2)^2} N(t-\gamma) dt. \]

Now,

\[ H(\gamma) = \int_{-\infty}^{\infty} \frac{1}{1+t^2} N(t-\gamma) dt, \]
thus

\begin{align*}
J'(\gamma) + H(\gamma) &= \int_{-\infty}^{\infty} \left[ \frac{1-t^2}{(1+t^2)^2} + \frac{1}{1+t^2} \right] N(t-\gamma) dt,
\end{align*}

\begin{align*}
&= 2 \int_{-\infty}^{\infty} \frac{1}{(1+t^2)^2} N(t-\gamma) dt,
\end{align*}

\begin{align*}
J'(\gamma) + H(\gamma) &= 2G(\gamma),
\end{align*}

as was to be shown.

The integral \( H(\gamma) \) has the series solution (9), and we now show that \( J'(\gamma) \) may be evaluated in terms of \( H(\gamma) \) and its derivatives with respect to \( \gamma \). Let

\begin{align*}
J(\gamma) &= \int_{-\infty}^{\infty} \frac{t}{1+t^2} N(t-\gamma) dt,
\end{align*}

\begin{align*}
H(\gamma) &= \int_{-\infty}^{\infty} \frac{1}{1+t^2} N(t-\gamma) dt,
\end{align*}

\begin{align*}
H'(\gamma) &= \int_{-\infty}^{\infty} \frac{t-\gamma}{1+t^2} N(t-\gamma) dt,
\end{align*}

\begin{align*}
H''(\gamma) &= \int_{-\infty}^{\infty} \frac{t^2-\gamma^2}{1+t^2} N(t-\gamma) dt.
\end{align*}
\[ J'(\gamma') = \int_{-\infty}^{\infty} \frac{t(t-\gamma')}{1+t^2} N(t-\gamma') \, dt, \]

then

\[ J(\gamma') = H'(\gamma') + \gamma H(\gamma'), \]

and

\[ J'(\gamma') = \int_{-\infty}^{\infty} \frac{t^2+1-\gamma t-1}{t^2+1} n(t-\gamma') \, dt = 1 - \gamma J(\gamma') - H(\gamma'), \]

so that

\[ 2G(\gamma') = J'(\gamma') + H(\gamma') = 1 - \gamma J(\gamma'). \]

Now,

\[ H(\gamma') = e^{\frac{-\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{\gamma^{2n} A_n}{n! 2^n}, \]

so that

\[ H'(\gamma') = -\frac{\gamma^2}{2} e^{\frac{-\gamma^2}{2}} \sum_{n=1}^{\infty} \frac{\gamma^{2n-1} A_n}{(n-1)! 2^{n-1}}, \]

\[ J(\gamma') = H'(\gamma') + \gamma H(\gamma') = e^{\frac{-\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{\gamma^{2n} A_{n+1}}{n! 2^n}, \]

\[ G(\gamma') = \frac{1}{2} e^{\frac{-\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{\gamma^{2n} A_{n+1}}{n! 2^n}. \]

Substituting the expressions (9) and (10) into (8),
\[
I(\gamma) = 1 + 3e^{-\frac{\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{2nA_n}{2^n n!} - \frac{5}{2} \left( 1 - e^{-\frac{\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{2nA_{n+1}}{2^n n!} \right),
\]

and algebraic simplification gives

\[
I(\gamma) = -\frac{3}{2} + e^{-\frac{\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{2nA_n}{2^n n!} \left[ 3A_n + 2.5 \gamma^2 A_{n+1} \right].
\]

Because the bracket in the last term contains \(\gamma^2\), it was decided that, in order to simplify the numerical evaluation of \(I(\gamma)\), a different form would be more desirable.

\[
2G(\gamma) = 1 - \frac{\gamma^2}{2}e^{-\frac{\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{2nA_n}{2^n n!}
\]

\[
= e^{-\frac{\gamma^2}{2}} \left[ \sum_{n=0}^{\infty} \frac{2nA_n}{2^n n!} - \sum_{n=0}^{\infty} \frac{2(n+1)A_{n+1}}{2^n (n+1)!} \right].
\]

\[
= e^{-\frac{\gamma^2}{2}} \left[ 1 + \sum_{n=1}^{\infty} \frac{2nA_n}{2^n n!} - \sum_{n=1}^{\infty} \frac{2n^2 A_n}{2^n n!} \right].
\]

\[
2G(\gamma) = e^{-\frac{\gamma^2}{2}} \left[ 1 + \sum_{n=1}^{\infty} \frac{2nA_n}{2^n n!} (1 - 2nA_n) \right].
\]

Then, since \(A_{n-1} = 1 - (2n-1)A_n\),

\[
G(\gamma) = e^{-\frac{\gamma^2}{2}} \left[ 1 + \sum_{n=1}^{\infty} \frac{2nA_n}{2^n n!} (A_{n-1} - A_n) \right].
\]
Using this last expression instead of (10) and simplifying, the integral $I(Y)$ becomes

$$I(Y) = 1 + e^{-\frac{Y^2}{2}} \left[ -2.5 + 3A_0 + \sum_{n=1}^{\infty} \frac{Y^{2n}}{2^n n!} \left( 5.5A_n - 2.5A_{n-1} \right) \right].$$

The mean square deviation in terms of powers of $Y$ is,

$$D^2_{Wo}(Y) = \sigma^2 \left\{ 2 + e^{-\frac{Y^2}{2}} \left[ -2.5 + 3A_0 + \sum_{n=1}^{\infty} \frac{Y^{2n}}{2^n n!} \left( 5.5A_n - 2.5A_{n-1} \right) \right] \right\}.$$

(11)

2. $E\left[ (W(t) - \mu)^2 \right]$ for the "sometimes pool" procedure, SP(t)

Mosteller (10) found the mean square deviation of SP(t) to be

$$D^2_{sp}(Y) = \sigma^2 \left\{ 1 + (t_\alpha + Y)N(t_\alpha + Y) + (t_\alpha - Y)N(t_\alpha - Y) + Y^2 \int_{-t_\alpha - Y}^{t_\alpha - Y} N(y)dy - t_\alpha - Y \right\} + \int_{-\infty}^{-t_\alpha - Y} N(y)dy + \int_{t_\alpha - Y}^{\infty} N(y)dy \right\},$$

(12)

where $Prob\left( |t| > t_\alpha \right) = a.$
E. Evaluation of the Bias

1. The bias of $W_0(t)$

The bias of $W(t)$ is given by the equation

$$B_W(\gamma) = \frac{\mu_1 - \mu_2}{2} - \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} t\phi(t)N(t-\gamma)dt.$$  

The bias, $B_{W_0}(\gamma)$, of $W_0(t)$ is obtained by substituting $\phi(t)$ for $\phi(t)$. This gives

$$B_{W_0}(\gamma) = \frac{\mu_1 - \mu_2}{2} - \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} \frac{t^3}{1+t^2} N(t-\gamma)dt.$$  

Making use of the fact that

$$\frac{t^3}{1+t^2} = t - \frac{t}{1+t^2},$$

the bias of $W_0(t)$ is

$$B_{W_0}(\gamma) = \frac{\mu_1 - \mu_2}{2} - \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} tN(t-\gamma)dt + \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} \frac{t}{1+t^2} N(t-\gamma)dt.$$

$$B_{W_0}(\gamma) = \frac{\mu_1 - \mu_2}{2} - \frac{\gamma \sigma}{\sqrt{2n}} + \frac{\sigma}{\sqrt{2n}} \int_{-\infty}^{\infty} \frac{t}{1+t^2} N(t-\gamma)dt.$$
By definition, see Equation (2),

\[ \frac{\gamma\sigma}{\sqrt{2\pi}} = \frac{\mu_1 - \mu_2}{Z} \]

and,

\[ J(\gamma) = \int_{-\infty}^{\infty} \frac{t}{1+t^2} N(t-\gamma) \, dt, \]

so that

\[ B_{\omega_0}(\gamma) = \frac{\sigma}{\sqrt{2\pi}} J(\gamma) = \frac{\gamma\sigma}{\sqrt{2\pi}} e^{-\frac{\gamma^2}{2}} \sum_{n=0}^{\infty} \frac{(2n\gamma A_{n+1})}{2^n n!}. \]  

(13)

2. The bias of SP(t)

To find the bias of the "sometimes pool" procedure, SP(t), we set

\[ \theta(t) = 0 \text{ if } |t| < t_a, \]
\[ \theta(t) = 1 \text{ if } |t| \geq t_a, \]

where \( t_a \) is the \( a/2 \) percentage point of the standard normal distribution. In this case the bias is

\[ B_{SP}(\gamma) = \frac{\mu_1 - \mu_2}{Z} - \frac{\sigma}{\sqrt{2\pi}} \left[ \int_{-\infty}^{-t_a} tN(t-\gamma) \, dt + \int_{t_a}^{\infty} tN(t-\gamma) \, dt \right]. \]
\[ z = \frac{\mu_1 - \mu_2}{2} - \frac{\sigma}{\sqrt{2\pi}} \left[ \int_{-\infty}^{t_a - \gamma} (z + \gamma) N(z) dz + \int_{t_a - \gamma}^{\infty} (z + \gamma) N(z) dz \right], \]

\[ B_{SP}(\gamma) = \frac{\mu_1 - \mu_2}{2} \cdot \frac{\gamma \sigma}{\sqrt{2\pi}} \left[ 1 - \int_{-\infty}^{t_a - \gamma} N(z) dz \right] + \frac{\sigma}{\sqrt{2\pi}} \left[ N(t_a + \gamma) - N(t_a - \gamma) \right], \]

\[ F, E_{\gamma} \]

\[ \begin{align*}
D_{W_0}^2(\gamma) &= \frac{\sigma^2}{\sqrt{2n}} \left\{ 2 + e^{-\gamma^2/2} \left[ 3A_0 - 2.5 + \sum_{n=1}^{\infty} \frac{\gamma^{2n}}{2^{n!} z^{2n}} (5.5A_n - 2.5A_{n-1}) \right] \right\}.
\end{align*} \]

The numerical values of \( D_{W_0}^2(\gamma) \) were calculated for \( \gamma \) equal to

\[ \frac{i - 1}{2}, \quad i = 1, 2, \ldots, 11, \]

and are tabulated in column 2 of Table 1 in units of \( \sigma^2/n \). Tables of the exponential function (5) were used to evaluate
Table 1. Mean square deviations in units of $\sigma^2/n$ and efficiencies relative to $X_1^2$ of $W_o(t)$ and $SP(t)$

<table>
<thead>
<tr>
<th>$Y$</th>
<th>$D_{W_o}^2(Y)$</th>
<th>$m(Y)$</th>
<th>R. E. ($W_o$)</th>
<th>$D_{SP}^2(Y)$</th>
<th>R. E. ($SP$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.734</td>
<td>0</td>
<td>1.363</td>
<td>0.732</td>
<td>1.366</td>
</tr>
<tr>
<td>0.5</td>
<td>0.780</td>
<td>2</td>
<td>1.282</td>
<td>0.849</td>
<td>1.178</td>
</tr>
<tr>
<td>1.0</td>
<td>0.892</td>
<td>4</td>
<td>1.121</td>
<td>1.118</td>
<td>0.895</td>
</tr>
<tr>
<td>1.5</td>
<td>1.011</td>
<td>5</td>
<td>0.989</td>
<td>1.362</td>
<td>0.734</td>
</tr>
<tr>
<td>2.0</td>
<td>1.091</td>
<td>7</td>
<td>0.917</td>
<td>1.394</td>
<td>0.717</td>
</tr>
<tr>
<td>2.5</td>
<td>1.113</td>
<td>9</td>
<td>0.898</td>
<td>1.363</td>
<td>0.734</td>
</tr>
<tr>
<td>3.0</td>
<td>1.122</td>
<td>11</td>
<td>0.891</td>
<td>1.218</td>
<td>0.821</td>
</tr>
<tr>
<td>3.5</td>
<td>1.107</td>
<td>13</td>
<td>0.903</td>
<td>1.099</td>
<td>0.910</td>
</tr>
<tr>
<td>4.0</td>
<td>1.089</td>
<td>16</td>
<td>0.919</td>
<td>1.035</td>
<td>0.967</td>
</tr>
<tr>
<td>4.5</td>
<td>1.072</td>
<td>19</td>
<td>0.933</td>
<td>1.010</td>
<td>0.991</td>
</tr>
<tr>
<td>5.0</td>
<td>1.059</td>
<td>23</td>
<td>0.944</td>
<td>1.002</td>
<td>0.998</td>
</tr>
</tbody>
</table>
In calculating \( D_{WO}^2 (\gamma) \) for a particular value of \( \gamma \), a sufficient number of terms in the series,

\[
\frac{-\gamma^2}{2} \sum_{n=1}^{\infty} e^{\frac{2nB_n}{z^n n!}},
\]

where

\[ B_n = 5.5A_n - 2.5A_{n-1} \]

was taken so that the remainder, in every case, would be less than .0005. The number of terms required, \( m(\gamma) \), for each \( \gamma \) for which \( D_{WO}^2 (\gamma) \) was evaluated, was found empirically and is shown in column 3 of Table 1.

The fourth column of Table 1 gives the efficiency of \( W_0(t) \) relative to \( X_1 \) for the corresponding value of \( \gamma \). The relative efficiency is

\[
\text{RE}(W_0) = \frac{\sigma^2}{nD_{WO}^2 (\gamma)}.
\]

The mean square deviation, \( D_{SP}^2 (\gamma) \), of the "sometimes pool" estimator, \( SP(t) \), is

\[
D_{SP}^2 (\gamma) = \frac{\sigma^2}{n}\left\{ 1 + \frac{t_a - \gamma}{N(t_a + \gamma)} + \frac{t_a - \gamma}{N(t_a - \gamma)} \right\}.
\]
This quantity, with \( t_a = 1.6 \), was evaluated for the range of \( \gamma \) given above and is recorded in units of \( \sigma^2/n \) in column 5 of Table 1. The critical value, \( t_a = 1.6 \), was selected so that \( W_o(t) \) and \( SP(t) \) would have very nearly the same efficiency for \( \gamma \) equal to zero. The corresponding relative efficiency,

\[
R.E.(SP) = \frac{\sigma^2}{nD_{SP}(\gamma)}
\]

is presented in the sixth column of Table 1. These calculations were performed with the aid of tables of ordinates and areas of the normal distribution function (8).

In Figure 1 the relative efficiencies of the two procedures under consideration are plotted as functions of \( \gamma \). Since the curves are symmetrical with respect to \( \gamma \), only the positive half of the \( \gamma \) axis is shown.

To illustrate the effect of sample size on the efficiency of \( W_o(t) \), the efficiency of \( W_o(t) \) relative to \( X_1 \) is plotted in Figure 2 as a function of \( |\mu_1 - \mu_2| \) for sample sizes \( n = 2, 8, 50, 200 \). The variance in each case is equal to one.

Figure 3 shows the efficiency of \( W_o(t) \) relative to \( X_1 \) as a function of \( |\mu_1 - \mu_2| \) for \( \sigma^2 = 25, 100, 225 \) with constant sample size, \( n = 50 \).

The bias of the weighted estimator,

\[
B_{W_o}(\gamma) = \frac{\sigma\gamma}{\sqrt{2n}} e^{-\frac{\gamma^2}{2}} \sum_{n=1}^{\infty} \frac{2n A_{n+1}}{2^n n!}
\]

was evaluated for
Figure 1. Efficiencies of \( W(H) \) and \( W(S) \)

Efficiency relative to \( \bar{X}_1 \)
Figure 2. Efficiency of $w_{0}(h)$ for different sample sizes when $z = 1$. 

1. $n = 50$
2. $n = 8$
3. $n = 2$
4. $n = 100$
Figure 3. Efficiency of $W^\circ (x)$ for different values of $u = 50$.
\[ \gamma = \frac{i - \frac{1}{2}}{2}, \quad i = 1, 2, \ldots, 11, \]

in units of \( \sigma / \sqrt{n} \) and is recorded in column 2 of Table 2. For each a sufficient number of terms of the series

\[ \sum_{n=0}^{\infty} \frac{z^n A_{n+1}}{2^n n!} \]

were taken so that the remainder would be less than .0005. The number of terms required, \( r(\gamma) \), for each value of \( \gamma \) for which the bias was evaluated, was determined empirically and is shown in column 3 of Table 2. Values of \( e^{-\gamma^2/2} \) were obtained from tables of the exponential function (5).

The bias of the "sometimes pool" estimator,

\[ B_{SP}(\gamma) = \frac{\sigma}{\sqrt{2\pi n}} \left[ \gamma \int_{-\gamma}^{\gamma} N(z)dz + N(t_a + \gamma) - N(t_a - \gamma) \right], \]

was evaluated for \( t_a = 1.6 \) and for the same range of \( \gamma \) as for \( B_W(\gamma) \). These values are given in units of \( \sigma / \sqrt{n} \) in column 4 of Table 2. They were calculated using tables of ordinates and areas of normal distribution function (8).

Figure 4 shows the biases of the two procedures as functions of \( \gamma \). The curves are given for only positive values of \( \gamma \) since
Table 2. Biases of $W_o(t)$ and $SP(t)$ in units of $\sigma/\sqrt{n}$

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$B_{W_o}(\gamma)$</th>
<th>$\tau(\gamma)$</th>
<th>$B_{SP}(\gamma)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.000</td>
<td>0</td>
<td>0.000</td>
</tr>
<tr>
<td>0.5</td>
<td>0.116</td>
<td>3</td>
<td>0.176</td>
</tr>
<tr>
<td>1.0</td>
<td>0.204</td>
<td>4</td>
<td>0.284</td>
</tr>
<tr>
<td>1.5</td>
<td>0.250</td>
<td>6</td>
<td>0.293</td>
</tr>
<tr>
<td>2.0</td>
<td>0.258</td>
<td>8</td>
<td>0.227</td>
</tr>
<tr>
<td>2.5</td>
<td>0.244</td>
<td>11</td>
<td>0.137</td>
</tr>
<tr>
<td>3.0</td>
<td>0.219</td>
<td>13</td>
<td>0.066</td>
</tr>
<tr>
<td>3.5</td>
<td>0.196</td>
<td>17</td>
<td>0.025</td>
</tr>
<tr>
<td>4.0</td>
<td>0.174</td>
<td>19</td>
<td>0.007</td>
</tr>
<tr>
<td>4.5</td>
<td>0.156</td>
<td>22</td>
<td>0.002</td>
</tr>
<tr>
<td>5.0</td>
<td>0.139</td>
<td>26</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Figure 4. Biases of $W_0(t)$ and $SP(t)$.
\[ B_{W_0}(-\gamma) = -B_{W_0}(\gamma) . \]

\[ B_{SP}(-\gamma) = -B_{SP}(\gamma) . \]

2. **Comparison of the efficiencies of \( W_0(t) \) and \( SP(t) \)**

A comparison of the relative efficiencies of \( W_0(t) \) and \( SP(t) \) tabulated in Table 1 and plotted in Figure 1 indicates that: (i) the maximum possible loss of efficiency is smaller for \( W_0(t) \) than for \( SP(t) \), the values of the maximum losses being .11 or 11 percent and .28 or 28 percent respectively; (ii) the largest value of \( |\gamma| \) for which the efficiency is not less than one is greater for \( W_0(t) \) than for \( SP(t) \), 1.4 for \( W_0(t) \), .8 for \( SP(t) \); (iii) for \( |\gamma| \) greater than 3.4, approximately, the efficiency of \( SP(t) \) is greater than that of \( W_0(t) \).

3. **Effect of sample size and variance on the efficiency of \( W_0(t) \)**

Let \( \xi \) be defined as the largest value of \( |\mu_1 - \mu_2| \) such that for all \( |\mu_1 - \mu_2| \leq \xi \) the relative efficiency of \( W_0(t) \) is greater than or equal to one. \( \xi \) will be referred to as the effective difference.

The effect of increasing sample size, holding the variance constant, on the effective difference is clearly shown by the curves of Figure 2. For \( n \) equal to eight the effective difference is approximately .7, but when \( n = 200 \) the effective difference is reduced to about .15. The larger the sample size the smaller the effective difference.
Figure 3 shows that for a fixed sample size, \( n = 50 \) for the curves shown, the effective difference increases as the variance increases; \( S = 1.4 \) for \( \sigma^2 = 25 \), but \( S = 4.3 \) for \( \sigma^2 = 225 \).

4. Comparison of the biases of \( W_\circ(t) \) and \( SP(t) \)

A comparison of the biases of \( W_\circ(t) \) and \( SP(t) \) as tabulated in Table 2 and plotted in Figure 4 shows that: (i) the absolute value of the maximum bias is less for \( W_\circ(t) \) than for \( SP(t) \); (ii) for \( |\gamma| \) less than or equal to approximately 1.8 the absolute value of the bias is smaller for \( W_\circ(t) \) than for \( SP(t) \); (iii) for \( |\gamma| \) greater than 1.8 the absolute value of the bias is smaller for \( SP(t) \) than for \( W_\circ(t) \).
IV. OTHER CASES OF POOLING MEANS

A. Different Known Variances

1. Statement of the problem

In the example of Section III it was assumed that equal sized samples were selected from two normal populations with a common known variance. Based on these assumptions a rule of procedure was set forth for pooling the sample means to form an estimator for the mean, \( \mu_1 \), of one of the populations. The mean square deviation of the resulting estimator was derived and used as a basis for comparing various estimators of \( \mu_1 \). This section considers the pooling of sample means from normal populations when the variances of the means are known but not equal. A rule of procedure is given below and the mean square deviation of the resulting estimator is derived.

2. Rule of procedure

Let \( \bar{X}_1 \) be the mean of a random sample taken from a normal population such that \( \bar{X}_1 \) is normally distributed with mean \( \mu_1 \) and variance \( V_1 \), and let \( \bar{X}_2 \) be the mean of a random sample from a second normal population such that \( \bar{X}_2 \) is normally distributed with mean \( \mu_2 \) and variance \( V_2 \), where \( V_1 \) and \( V_2 \) are known. Under these
assumptions an appropriate statistic for testing the hypothesis that
\[ \mu_1 = \mu_2 \] is the normal deviate

\[ t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{1}{V_1} + \frac{1}{V_2}}} \]  \hspace{1cm} (15)

which has expectation

\[ \gamma = \frac{\mu_1 - \mu_2}{\sqrt{\frac{1}{V_1} + \frac{1}{V_2}}} \]  \hspace{1cm} (16)

To estimate \( \mu_1 \) use

\[ W(t) = \phi(t) \overline{X}_1 + \left[1 - \phi(t)\right] \frac{V_2 \overline{X}_1 + V_1 \overline{X}_2}{\sqrt{V_1 + V_2}} \]  \hspace{1cm} (17)

where \( \phi(t) \) is a function of \( t \) only.

3. Derivation of the mean square deviation of \( W(t) \)

Rewrite \( W(t) \) as

\[ W(t) = \frac{V_2 \overline{X}_1 + V_1 \overline{X}_2}{\sqrt{V_1 + V_2}} + \frac{t\phi(t)V_1}{\sqrt{V_1 + V_2}} \]

The mean square deviation of \( W(t) \) is

\[ D_{W}^2(\gamma) = \mathbb{E} \left[ \frac{V_2 \overline{X}_1 + V_1 \overline{X}_2}{\sqrt{V_1 + V_2}} + \frac{t\phi(t)V_1}{\sqrt{V_1 + V_2}} - \mu_1 \right]^2 \]

or
\[ D_W^2(\chi) = E \left[ \frac{V_2X_1 + V_1X_2}{V_1 + V_2} \right]^2 + \frac{V_1^2}{V_1 + V_2} E \left[ t\theta(t) \right]^2 + \mu_1^2 \]

\[ + \frac{2V_1}{\sqrt{V_1 + V_2}} E \left[ \frac{V_2X_1 + V_1X_2}{V_1 + V_2} \right] t\theta(t) - 2\mu_1 E \left[ \frac{V_2X_1 + V_1X_2}{V_1 + V_2} \right] \]

\[ - \frac{2\mu_1 V_1}{\sqrt{V_1 + V_2}} E \left[ t\theta(t) \right]. \]

Substituting the expected values and making use of the fact that

t and \( (V_2X_1 + V_1X_2) / (V_1 + V_2) \) are independent,

\[ D_W^2(\chi) = \frac{V_1 V_2}{V_1 + V_2} + \left[ \frac{V_2\mu_1 + V_1\mu_2}{V_1 + V_2} \right]^2 + \mu_1 - 2\mu_1 \left[ \frac{V_2\mu_1 + V_1\mu_2}{V_1 + V_2} \right] \]

\[ + \frac{V_1}{V_1 + V_2} \int_{-\infty}^{\infty} t^2 \theta^2(t) N(t - \chi) dt + \frac{2V_1}{\sqrt{V_1 + V_2}} \left[ \frac{V_2\mu_1 + V_1\mu_2}{V_1 + V_2} \right] \int_{-\infty}^{\infty} t\theta(t) N(t - \chi) dt \]

\[ - \frac{2\mu_1 V_1}{\sqrt{V_1 + V_2}} \int_{-\infty}^{\infty} t\theta(t) N(t - \chi) dt. \]

Algebraic simplification yields

\[ D_W^2(\chi) = \frac{V_1 V_2}{V_1 + V_2} + \frac{V_1^2 (\mu_1 - \mu_2)^2}{(V_1 + V_2)^2} + \frac{V_1^2}{V_1 + V_2} \int_{-\infty}^{\infty} t^2 \theta(t) N(t - \chi) dt \]

\[ - \frac{2V_1 (\mu_1 - \mu_2)}{(V_1 + V_2)^{3/2}} \int_{-\infty}^{\infty} t\theta(t) N(t - \chi) dt. \]
or, since \( \gamma = \frac{(\mu_1 - \mu_2)}{\sqrt{\nu_1 + \nu_2}} \)

\[
D_w^2(\gamma) = \frac{V_1 V_2}{\nu_1 + \nu_2} + \frac{V_1^2 \gamma^2}{\nu_1 + \nu_2} + \frac{V_1^2}{\nu_1 + \nu_2} \int_{-\infty}^{\infty} t^2 \varphi^2(t) N(t - \gamma) dt
\]

\[
- \frac{2V_1^2 \gamma}{\nu_1 + \nu_2} \int_{-\infty}^{\infty} t \varphi(t) N(t - \gamma) dt.
\]

Factoring the right hand member,

\[
D_w^2(\gamma) = \frac{V_1}{\nu_1 + \nu_2} \left\{ V_2 + V_1 \int_{-\infty}^{\infty} \left[ t \varphi(t) - \gamma \right]^2 N(t - \gamma) dt \right\},
\]

or

\[
D_w^2(\gamma) = \frac{V_1}{1 + \frac{V_1}{\nu_2}} \left\{ 1 + \frac{V_1}{\nu_2} \int_{-\infty}^{\infty} \left[ t \varphi(t) - \gamma \right]^2 N(t - \gamma) dt \right\}.
\]

(18)
B. Some Special Cases

1. Equal population variances, different sample sizes

Let \( \bar{X}_1 \) and \( \bar{X}_2 \) be the means of independent random samples of size \( n_1 \) and \( n_2 \) respectively taken from two normal populations with a common known variance, \( \sigma^2 \), and means \( \mu_1 \) and \( \mu_2 \). The variance of \( \bar{X}_1 \) is

\[
\sigma^2_{\bar{X}_1} = \frac{\sigma^2}{n_1},
\]

and the variance of \( \bar{X}_2 \) is

\[
\sigma^2_{\bar{X}_2} = \frac{\sigma^2}{n_2}.
\]

If these variances are substituted in Equation (18) above, the mean square deviation of \( W(t) \) is

\[
D_w^2(\gamma) = \sigma^2 \left\{ 1 + \frac{n_2}{n_1} \int_{-\infty}^{\infty} \left[ t \phi(t) - \gamma \right]^2 N(t-\gamma) dt \right\}. \tag{19}
\]

2. Different population variances, equal sample sizes

If \( \bar{X}_1 \) and \( \bar{X}_2 \) are based on equal sample sizes, \( n_1 = n_2 = n \), and if the variances of the populations concerned are \( \sigma^2_1 \) and \( \sigma^2_2 \), then

\[
\sigma^2_{\bar{X}_1} = \frac{\sigma^2_1}{n},
\]

\[
\sigma^2_{\bar{X}_2} = \frac{\sigma^2_2}{n},
\]

and

\[ D_{W}^{2}(\bar{y}) = \frac{\sigma_{1}^{2}}{\sigma_{1}^{2} + \sigma_{2}^{2}} \left\{ 1 + \frac{\sigma_{1}^{2}}{\sigma_{2}^{2}} \int_{-\infty}^{\infty} \left[ t\delta(t) - \delta \right]^{2} N(t - \bar{y}) dt \right\}. \tag{20} \]

3. Different population variances, different sample sizes

If \( \bar{X}_{1} \) is the mean of a sample of size \( n_{1} \) selected from a normal population with mean \( \mu_{1} \) and variance \( \sigma_{1}^{2} \), and if \( \bar{X}_{2} \) is the mean of a sample of size \( n_{2} \) selected from a normal population with mean \( \mu_{2} \) and variance \( \sigma_{2}^{2} \), then

\[ \sigma_{\bar{X}_{1}}^{2} = \frac{\sigma_{1}^{2}}{n_{1}}, \]
\[ \sigma_{\bar{X}_{2}}^{2} = \frac{\sigma_{2}^{2}}{n_{2}}. \]

and

\[ D_{W}^{2}(\bar{y}) = \frac{\sigma_{1}^{2} n_{2}^{2}}{n_{1}^{2} \left[ 1 + \frac{n_{2}^{2} \sigma_{1}^{2}}{n_{1}^{2} \sigma_{2}^{2}} \right]} \left\{ 1 + \frac{n_{2}^{2} \sigma_{1}^{2}}{n_{1}^{2} \sigma_{2}^{2}} \int_{-\infty}^{\infty} \left[ t\delta(t) - \delta \right]^{2} N(t - \bar{y}) dt \right\}. \tag{21} \]
C. Results

The mean square deviation of $W(t)$, Equation (18), in addition to being a function of the nuisance parameter, $\gamma$, is also a function of the ratio, $V_1/V_2$, of the variances of the sample means. In order to investigate the effect of $V_1/V_2$, the mean square deviation obtained in Section III for the special case, $V_1/V_2 = 1$, will be used as a basis for comparison.

Let $\gamma_1$ be that finite value of $|\gamma|$ such that, for $\gamma = \pm \gamma_1$, the integral

$$K(\gamma) = \int_{-\infty}^{\infty} \left[ t\phi(t) - \gamma \right]^2 N(t-\gamma) \, dt,$$

is equal to one. In general, $\gamma_1$ will have a different numerical value for each different function $\phi(t)$. If $\phi(t)$ is defined to be

$$\phi_o(t) = \frac{t^2}{1+t^2},$$

then $\gamma_1$ is approximately equal to 1.45. This figure was determined empirically. If $|\gamma|$ is less than $\gamma_1$ the integral, $K(\gamma)$, will be less than one. If $|\gamma|$ is greater than one, $K(\gamma)$ is greater than one, except for $|\gamma| = \infty$ which makes $K(\gamma)$ equal to one.

If Equation (18) is rewritten as

$$D_W^2(\gamma) = \frac{V_1 \left[ 1 + \frac{V_1}{V_2} K(\gamma) \right]}{1 + \frac{V_1}{V_2}},$$
it is readily seen that when the ratio, $V_1/V_2$, is less than one; (1) for all $|\gamma| < \gamma_1$ the mean square deviation will be greater than that of the special case, $V_1/V_2 = 1$, and (2) for all $|\gamma| > \gamma_1$ the mean square deviation will be less than when $V_1/V_2 = 1$. If $V_1/V_2$ is greater than one; (1) for all $|\gamma| < \gamma_1$ the mean square deviation is less than when $V_1/V_2 = 1$, and (2) for all $|\gamma| > \gamma_1$ the mean square deviation is greater than for $V_1/V_2 = 1$.

The ratio $V_1/V_2$ is less than one when

(i) $n_1 > n_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

(ii) $n_1 = n_2 = n$ and $\sigma_1^2 < \sigma_2^2$,

(iii) $n_2\sigma_1^2 < n_1\sigma_2^2$,

and is greater than one when

(i) $n_1 < n_2$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$,

(ii) $n_1 = n_2 = n$ and $\sigma_1^2 > \sigma_2^2$,

(iii) $n_2\sigma_1^2 > n_1\sigma_2^2$.
V. POOLING NORMAL OR NEAR NORMAL ESTIMATORS

A. Normally Distributed Estimators

In Sections III and IV the generalized pooling procedure of Section II was considered in connection with pooling sample means from normal distributions. This procedure will offer the same advantages over the preliminary testing procedure when pooling other normally distributed estimators. Let \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \) be two unbiased, normally and independently distributed estimators for the parameters \( \theta_1 \) and \( \theta_2 \) respectively. Let the variance of \( \hat{\theta}_1 \) be \( \sigma_1^2 \), the variance of \( \hat{\theta}_2 \) be \( \sigma_2^2 \), and let \( \sigma_1^2 \) and \( \sigma_2^2 \) be known. Then, if there is reason to believe that the difference between \( \theta_1 \) and \( \theta_2 \) is small or equal to zero, a pooled estimator for \( \theta_1 \) is obtained from the weighting procedure.

To estimate \( \theta_1 \) use

\[
W(t, \hat{\theta}_1, \hat{\theta}_2) = \phi(t)\hat{\theta}_1 + [1 - \phi(t)] \frac{\sigma_2^2\hat{\theta}_1 + \sigma_1^2\hat{\theta}_2}{\sigma_1^2 + \sigma_2^2},
\]

(22)

where \( \phi(t) \) is a function of \( t \) only and

\[
t = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}},
\]
is a normal deviate with variance equal to one. The expectation of
\( t \) is
\[
\gamma = \frac{\hat{\theta}_1 - \hat{\theta}_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}.
\]

The mean square deviation of \( W(t, \hat{\theta}_1, \hat{\theta}_2) \) is the same as that
for the case of pooling sample means with different known variances,
that is,
\[
D^2_W(\gamma) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \left\{ \int_{-\infty}^{\infty} \left[ t\theta(t) - \gamma \right]^2 N(t - \gamma) dt \right\}.
\]
(23)

If the efficiency of \( W(t, \hat{\theta}_1, \hat{\theta}_2) \) is plotted as a function of \( \gamma \) when
\( \sigma_1^2 = \sigma_2^2 \), and when \( \theta(t) = t^2/(1+t^2) \), the figure obtained is identical
with the curve in Figure 1 which shows the efficiency of \( W_0(t) \) relative
to \( \bar{X}_1 \) as an estimator of \( \mu_1 \).

The expected value of \( W(t, \hat{\theta}_1, \hat{\theta}_2) \) is
\[
E\left[ W(t, \hat{\theta}_1, \hat{\theta}_2) \right] = E\left[ \frac{\sigma_1^2 \hat{\theta}_1 + \sigma_2^2 \hat{\theta}_2}{\sigma_1^2 + \sigma_2^2} \right] + \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} E\left[ t\theta(t) \right].
\]

\[
E\left[ W(t, \hat{\theta}_1, \hat{\theta}_2) \right] = \frac{\sigma_1^2 \hat{\theta}_1 + \sigma_2^2 \hat{\theta}_2}{\sigma_1^2 + \sigma_2^2} + \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{-\infty}^{\infty} t\theta(t)N(t - \gamma) dt.
\]

The bias of \( W(t, \hat{\theta}_1, \hat{\theta}_2) \) as an estimator for \( \theta_1 \) is
B. Approximately Normal Estimators

In the practical applications of Statistics it sometimes happens that the estimators have sampling distributions which are not known but which can be assumed to be approximately normal for sufficiently large sample sizes. In other cases the distribution of the estimator is known but is so very complex that it is cumbersome to deal with in practical work. In some of these cases a transformation is used which yields a new variables which is approximately normally distributed for sufficiently large samples. A third situation of interest is one which occurs when an estimator is transformed so as to make the variance of the new variable independent of the parameter to be estimated. Some transformed variables of this type are approximately normally distributed. In any of the aforementioned cases, if two independent estimators are available and if pooling of the estimators is considered, the generalized pooling procedure is applicable and, if the approximation to a normal distribution is good, the formulas which have been derived for the mean square deviation and the bias should be approximately equal to the true mean square deviation and
bias of the pooled estimator. Examples of variables which are approximately normally distributed may be found in most texts on statistical methods or the theory of statistics. Hald (7), in the index of his book on statistical theory, lists normal approximations for the following: chi-square, coefficient of variation, estimate of $\sigma$ in truncated distributions, mean, number of runs, mean square successive difference, relative frequency, standard deviation, "student's" distribution, transformed correlation coefficient, transformed relative frequency, and variance.

A pooling procedure is not of interest in all of these cases, nor is the approximation equally good for all cases. One of these approximations, the transformed correlation coefficient, is considered in some detail in Section VIII.
VI. WEIGHTING FUNCTIONS

If the rule of procedure of Section V is to be used for pooling two independent normal estimators, \( \hat{\theta}_1 \) and \( \hat{\theta}_2 \), in order to estimate the parameter \( \theta_1 \), the weighting function, \( \phi(t) \), must first be selected. The choice of \( \phi(t) \) will be restricted to the class of single valued functions of \( t \) which are continuous except on a set of measure zero, which exist for all \( t \), and which satisfy the conditions

(i) \[ 0 \leq \phi(t) \leq 1, \text{ for all } t, \]

(ii) \( \phi(-t) = \phi(t) \).

The class of functions so defined will be referred to as the admissible class of weighting functions.

The choice of a weighting function should be based on some criterion by means of which the relative merits of various alternative functions may be assessed. A possible criterion is unbiasedness; that is, if a function \( \phi_u(t) \) exists such that, for all \( \gamma \), the expected value of

\[
W_u(t) = \phi_u(t) \hat{\theta}_1 + \frac{1 - \phi_u(t)}{\sigma_1^2 \phi_1 + \sigma_2^2 \phi_2} \cdot \frac{\sigma_1^2 \hat{\theta}_1 + \sigma_2^2 \hat{\theta}_2}{\sigma_1^2 + \sigma_2^2},
\]

is equal to \( \theta_1 \), then \( \phi_u(t) \) is an unbiased weighting function.

A second desirable property is uniformly minimum variance about \( \theta_1 \). If \( \phi_v(t) \) is a weighting function such that for all \( \gamma \) the mean square deviation of \( W_{\phi_v}(t) \) is less than or equal to the mean square deviation of
where $\phi'(t)$ is any other admissible weighting function, and if the inequality holds for some range of $x$, then $\phi_y(t)$ is a uniformly minimum variance weighting function. It will be shown that among the class of admissible functions the only unbiased weighting function is $\phi(t) = 1$. It will also be shown that there does not exist, among the class of admissible functions, a uniformly minimum variance weighting function.

A third criterion which may be proposed is one which leads to the choice of the weighting function which yields an estimator whose efficiency is greatest when averaged, in some sense, for all values of $x$. One such measure of overall efficiency is the area between the curves corresponding to the mean square deviation of $W(t)$ and to the variance of the never pool estimator $\hat{\theta}_1$. The area is

$$I = \int_{-\infty}^{\infty} \left[ D_{NP}(x) - D_{W}(x) \right] \, dy,$$

and the best weighting function, according to this criterion, is the one for which the integral, $I$, is a maximum. It will be shown that the function which maximizes the integral is $\phi(t) = 1$, for which $W(t) = \hat{\theta}_1$.

Theorem 1: Among the class of admissible weighting functions the only unbiased weighting function is $\phi(t) = 1$.

Proof: The bias of $W(t)$ as an estimator of $\theta_1$ is, from equation (24),
\[ B_W(\gamma) = \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \left[ \gamma - \int_{-\infty}^{\infty} t\theta(t)N(t-\gamma)dt \right]. \]

Since
\[ \gamma = E(t) = \int_{-\infty}^{\infty} tN(t-\gamma)dt, \]
the bias is also given by
\[ B_W(\gamma) = \frac{\sigma_1^2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \int_{-\infty}^{\infty} t \left[ 1 - \theta(t) \right] N(t-\gamma)dt. \]

If a function \( \theta(t) \) exists such that \( B_W(\gamma) = 0 \) for all \( \gamma \), it must satisfy the condition that
\[ \int_{-\infty}^{\infty} t \left[ 1 - \theta(t) \right] N(t-\gamma)dt = 0, \]
or that
\[ \int_{-\infty}^{0} t \left[ 1 - \theta(t) \right] N(t-\gamma)dt + \int_{0}^{\infty} t \left[ 1 - \theta(t) \right] N(t-\gamma)dt = 0, \]
or, if \( t \) is replaced by \(-t\) in the first integral,
\[ -\int_{0}^{\infty} t \left[ 1 - \theta(-t) \right] N(t+\gamma)dt + \int_{0}^{\infty} t \left[ 1 - \theta(t) \right] N(t-\gamma)dt = 0. \]
If \( \tilde{\phi}(t) \) is an admissible weighting function it satisfies the condition that 
\( \tilde{\phi}(-t) = \tilde{\phi}(t) \). Therefore the condition that \( \tilde{\phi}(t) \) must satisfy, if it is to give an unbiased estimator when \( \gamma \neq 0 \), is

\[
\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} t \left[ 1 - \tilde{\phi}(t) \right] \left[ e^{-\frac{1}{2}(t-\gamma)^2} - e^{-\frac{1}{2}(t+\gamma)^2} \right] dt = 0.
\]

For any value of \( \gamma \) not equal to zero and for \( 0 < t < \infty \), the factor

\[
\frac{1}{2}(t-\gamma)^2 - e^{-\frac{1}{2}(t+\gamma)^2},
\]

is either greater than zero for all \( t \) or less than zero for all \( t \) depending upon whether \( \gamma \) is a positive or negative quantity. It follows that the integral can be equal to zero when \( \gamma \) is not equal to zero if and only if \( \tilde{\phi}(t) \) is identically equal to one, as was to be proved.

Theorem 2: Among the class of admissible weighting functions, that one which maximizes the integral

\[
I = \int_{-\infty}^{\infty} \left[ D_{NP}(\gamma) - D_{W}(\gamma) \right] d\gamma
\]

is \( \tilde{\phi}(t) = 1 \).

Proof: From Equation (23), the mean square deviation of \( W(t) \) is

\[
D_{W}(\gamma) = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \left\{ \sigma_2^2 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} \int_{-\infty}^{\infty} \left[ t\tilde{\phi}(t) - \gamma \right]^2 N(t-\gamma)dt \right\},
\]

the variance of \( \hat{\theta}_1 \) is
\[ D_{\text{NE}}^2(\gamma) = \sigma_1^2, \]

so that the integral, I, is

\[ I = \frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2} \int_{-\infty}^{\infty} \left\{ 1 - \int_{-\infty}^{\infty} \left[ t\phi(t) - \delta \right]^2 N(t, \gamma) \, dt \right\} \, d\gamma, \]

or

\[ I = \frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ 1 - \left[ t\phi(t) - \delta \right]^2 \right\} N(t, \gamma) \, dt \, d\gamma. \]

Integrating with respect to \( \gamma \),

\[ I = -\frac{\sigma_1^4}{\sigma_1^2 + \sigma_2^2} \int_{-\infty}^{\infty} t^2 \left[ 1 - \phi(t) \right]^2 \, dt. \]

The integrand is always positive unless \( \phi(t) = 1 \). The integral is therefore always negative except for \( \phi(t) = 1 \). Thus the maximum value for the integral, I, is equal to zero and this maximum is attained if and only if \( \phi(t) = 1 \).

Theorem 3: Among the class of admissible weighting functions there exists no function \( \phi_v(t) \) such that

\[ D_{\phi_v}^2(\gamma) \leq D_{\phi_1}^2(\gamma), \]

for all \( \gamma \) and for any other admissible function \( \phi'(t) \).
Proof: As a consequence of Theorem 2, the maximum value of the integral,

\[ I = \int_{-\infty}^{\infty} \left[ D_{\text{NP}}(\lambda) - D_{W}(\lambda) \right] d\lambda, \]

is equal to zero. Therefore there is no weighting function such that the mean square deviation of \( W(t) \) is less than the variance of \( \hat{\theta}_1 \) for all \( \lambda \), for if such a function did exist the integral, \( I \), would be positive. This implies that for every admissible function not equal to one almost everywhere there is some range of \( \lambda \) for which the mean square deviation of \( W(t) \) must be greater than the variance of \( \hat{\theta}_1 \). Furthermore, \( \varnothing(t) = 1 \) is not a uniformly minimum variance weighting function for when \( \gamma \) is equal to zero the variance of the "always pool" estimator,

\[ \frac{\sigma_2^2 \hat{\theta}_1 + \sigma_1^2 \hat{\theta}_2}{\sigma_1^2 + \sigma_2^2} \]

is

\[ D_{\text{AP}}(\lambda) = \frac{\sigma_1^2}{1 + \frac{\sigma_1^2}{\sigma_2^2}} < \sigma_1^2. \]

In the example of Section III a weighting function

\[ \varnothing_0(t) = \frac{t^2}{1 + t^2} \]

was used because \( t \) is an unbiased estimator for \( \gamma \), and because the
minimum mean square deviation of \( W(t) \) for a given value of \( \gamma \) is obtained when

\[
\Phi(t) = \frac{\gamma^2}{1 + \gamma^2},
\]

for all \( t \). This is also true when pooling normal estimators with different known variances. In addition to satisfying the restrictions on the class of admissible weighting functions, \( \Phi_o(t) \) is a monotone increasing function of \( |t| \). The monotone property seems intuitively to be a desirable one.

The function \( \Phi_o(t) \) does not involve parameters which can be varied to study the effects of changing the shape of the curve associated with the weighting function on the mean square deviation of \( W(t) \). For this reason a two-parameter family of weighting functions is introduced in Section VII and the effects of varying the parameters are investigated.
VII. A TWO-PARAMETER FAMILY OF WEIGHTING FUNCTIONS

A. The Family Defined

A two-parameter family of weighting functions is defined by

\[ \phi(t; a, b) = 1 - ae^{-bt^2} \]  \hspace{1cm} (26)

where \( a \) and \( b \) are parameters such that

(i) \( 0 \leq a \leq 1 \),

(ii) \( b \geq 0 \).

The restrictions on \( a \) and \( b \) are necessary to insure that

\[ 0 \leq \phi(t; a, b) \leq 1, \]

for all \( t \).

B. The Mean Square Deviation of \( W_{ab}(t) \)

Returning to the special case of Section III, pooling sample means with known common variance and equal sample sizes, the estimator for \( \mu_1 \) obtained by using \( \phi(t; a, b) \) is

\[ W_{ab}(t) = \frac{\bar{X}_1 + \bar{X}_2}{2} + \phi(t; a, b) \frac{\bar{X}_1 - \bar{X}_2}{2}, \]

and the mean square deviation of \( W_{ab}(t) \) is
\[
D_{ab}^{2}(\gamma) = \frac{e^{2}}{2\pi} \left\{ 1 + \int_{-\infty}^{\infty} \left[ t\phi(t; a, b) - \gamma \right]^{2} N(t-\gamma)dt \right\} ,
\]
or, substituting from Equation (26),
\[
D_{ab}^{2}(\gamma) = \frac{e^{2}}{2\pi} \left\{ 1 + \int_{-\infty}^{\infty} \left[ -t(1 - ae^{-bt^{2}}) \right]^{2} N(t-\gamma)dt \right\} .
\]
Rearranging the integrand, the integral,
\[
G(\gamma) = \int_{-\infty}^{\infty} \left[ \gamma - t + at e^{-bt^{2}} \right]^{2} N(t-\gamma)dt,
\]
becomes
\[
G(\gamma) = \int_{-\infty}^{\infty} (t-\gamma)^{2} N(t-\gamma)dt - 2a \int_{-\infty}^{\infty} (t-\gamma)te^{-bt^{2}} N(t-\gamma)dt
\]
\[
+ a^{2} \int_{-\infty}^{\infty} t^{2} e^{-2bt^{2}} N(t-\gamma)dt . \tag{27}
\]

The second integral of Equation (27) is
\[
G_{2}(\gamma) = \int_{-\infty}^{\infty} (t-\gamma)te^{-bt^{2}} N(t-\gamma)dt = \int_{-\infty}^{\infty} \frac{t(t-\gamma)}{\sqrt{2\pi}} e^{-\frac{1}{2} (t-\gamma)^{2}-bt^{2}} dt.
\]
The transformation,

\[ y = \sqrt{2b+1} \left( t - \frac{\chi}{2b+1} \right), \quad dy = \sqrt{2b+1} \ dt, \]

leads to

\[
G_2(\gamma) = e^{\frac{b \gamma^2}{2b+1}} \int_{-\infty}^{\infty} \left( \frac{t - \gamma}{\sqrt{2b+1}} \right) e^{-\frac{1}{2} \left( \frac{2b+1}{2} t^2 - 2b \gamma^2 \right)} \ dt,
\]

\[
G_2(\gamma) = e^{\frac{b \gamma^2}{2b+1}} \int_{-\infty}^{\infty} \left( \frac{t - \gamma}{\sqrt{2b+1}} \right) e^{-\frac{1}{2} \left( \frac{2b+1}{2} t^2 - 2b \gamma^2 \right)} \ dt.
\]

The third integral of Equation (27) is integrated in a similar manner.

\[
G_3(\gamma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left( t - \gamma \right)^2 - 2b t^2} dt.
\]
Using the transformation,

\[ X = \sqrt{4b+1} \left( t - \frac{\gamma}{4b+1} \right), \quad dx = \sqrt{4b+1} \, dt, \]

the integral becomes

\[
G_3(\gamma) = \frac{1}{\sqrt{2\pi}} e^{-\frac{2b \gamma^2}{4b+1}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2} \left( \sqrt{4b+1} \left( t - \frac{\gamma}{4b+1} \right) \right)^2} \, dt.
\]

Combining the results shown in Equations (28) and (29) with Equation (27), the mean square deviation of \( W_{ab}(t) \) is

\[
D_{ab}^2(\gamma) = \frac{\sigma^2}{2n} \left\{ 2 - \frac{2a}{(2b+1)^{3/2}} \left[ 1 - \frac{2b \gamma^2}{2b+1} \right] e^{-\frac{b \gamma^2}{2b+1}} \right. \\
+ \left. \frac{a^2}{(4b+1)^{3/2}} \left[ 1 + \frac{\gamma^2}{4b+1} \right] e^{-\frac{2b \gamma^2}{4b+1}} \right\}.
\]
C. Results

1. Tables

The mean square deviation of $W_{ab}(t)$ is recorded in Table 3 for various combinations of $a$ and $b$, for values of $|Y|$ from zero to six in steps of one-half, and for $|Y|$ equal to seven. The entries in the table are in units of $\sigma^2/n$. The efficiency of $W_{ab}(t)$, relative to $X_1$, as an estimator for $\mu_1$ is given in Table 4, for the same values of $|Y|$ and for the same combinations of $a$ and $b$ as the mean square deviation. The calculations were made with the aid of tables of the exponential function (5).

2. Effects of varying $b$ when $a$ is held constant

If the parameter $a$ is specified as being equal to some value greater than zero but not greater than one, and if the parameter $b$ is then allowed to assume different values, $b > 0$, it is possible to study the effect of the selection of $b$ on the efficiency of the estimator $W_{ab}(t)$. In Figure 5 the relative efficiency curves are plotted as functions of $|Y|$ for $a = 1.00$ and $b = 1.00, .50, .25, .10$. These curves reveal that decreasing $b$ when $a$ is constant has the following effects: (1) the efficiency when $|Y|$ is equal to zero is increased, (2) the maximum possible loss of efficiency is increased, and (3) the range of $|Y|$ for which large losses of efficiency may be sustained is increased without a corresponding increase in the effective difference.
Table 3. The mean square deviation of $W_{ab}(t)$ in units of $\sigma^2/n$

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>a = 1.00</th>
<th>a = 1.00</th>
<th>a = 1.00</th>
<th>a = 1.00</th>
<th>a = .65</th>
<th>a = .42</th>
</tr>
</thead>
<tbody>
<tr>
<td>b = 1.00</td>
<td>b = .50</td>
<td>b = .25</td>
<td>b = .10</td>
<td>b = .10</td>
<td>b = .10</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>.852</td>
<td>.743</td>
<td>.732</td>
<td>.541</td>
<td>.633</td>
<td>.734</td>
</tr>
<tr>
<td>0.5</td>
<td>.940</td>
<td>.805</td>
<td>.708</td>
<td>.629</td>
<td>.681</td>
<td>.761</td>
</tr>
<tr>
<td>1.0</td>
<td>.990</td>
<td>.954</td>
<td>.899</td>
<td>.865</td>
<td>.810</td>
<td>.834</td>
</tr>
<tr>
<td>1.5</td>
<td>1.072</td>
<td>1.105</td>
<td>1.120</td>
<td>1.176</td>
<td>.985</td>
<td>.935</td>
</tr>
<tr>
<td>2.0</td>
<td>1.101</td>
<td>1.190</td>
<td>1.288</td>
<td>1.476</td>
<td>1.160</td>
<td>1.040</td>
</tr>
<tr>
<td>2.5</td>
<td>1.084</td>
<td>1.195</td>
<td>1.360</td>
<td>1.694</td>
<td>1.298</td>
<td>1.127</td>
</tr>
<tr>
<td>3.0</td>
<td>1.051</td>
<td>1.150</td>
<td>1.345</td>
<td>1.799</td>
<td>1.379</td>
<td>1.185</td>
</tr>
<tr>
<td>3.5</td>
<td>1.024</td>
<td>1.093</td>
<td>1.277</td>
<td>1.797</td>
<td>1.402</td>
<td>1.210</td>
</tr>
<tr>
<td>4.0</td>
<td>1.010</td>
<td>1.048</td>
<td>1.193</td>
<td>1.716</td>
<td>1.378</td>
<td>1.208</td>
</tr>
<tr>
<td>4.5</td>
<td>1.003</td>
<td>1.022</td>
<td>1.120</td>
<td>1.593</td>
<td>1.327</td>
<td>1.186</td>
</tr>
<tr>
<td>5.0</td>
<td>1.001</td>
<td>1.008</td>
<td>1.066</td>
<td>1.460</td>
<td>1.263</td>
<td>1.154</td>
</tr>
<tr>
<td>5.5</td>
<td>1.000</td>
<td>1.002</td>
<td>1.034</td>
<td>1.338</td>
<td>1.199</td>
<td>1.119</td>
</tr>
<tr>
<td>6.0</td>
<td>1.000</td>
<td>1.001</td>
<td>1.015</td>
<td>1.236</td>
<td>1.143</td>
<td>1.088</td>
</tr>
<tr>
<td>7.0</td>
<td>1.000</td>
<td>1.000</td>
<td>1.002</td>
<td>1.102</td>
<td>1.064</td>
<td>1.040</td>
</tr>
</tbody>
</table>
Table 4. Efficiency of $W_{ab}(t)$ relative to $X_1$

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$b = 1.00$</th>
<th>$b = 0.50$</th>
<th>$b = 0.25$</th>
<th>$b = 0.10$</th>
<th>$b = 0.10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>1.173</td>
<td>1.346</td>
<td>1.581</td>
<td>1.848</td>
<td>1.580</td>
</tr>
<tr>
<td>0.5</td>
<td>1.064</td>
<td>1.242</td>
<td>1.412</td>
<td>1.589</td>
<td>1.469</td>
</tr>
<tr>
<td>1.0</td>
<td>1.010</td>
<td>1.048</td>
<td>1.112</td>
<td>1.156</td>
<td>1.234</td>
</tr>
<tr>
<td>1.5</td>
<td>0.933</td>
<td>0.905</td>
<td>0.892</td>
<td>0.850</td>
<td>1.015</td>
</tr>
<tr>
<td>2.0</td>
<td>0.908</td>
<td>0.841</td>
<td>0.776</td>
<td>0.678</td>
<td>0.862</td>
</tr>
<tr>
<td>2.5</td>
<td>0.922</td>
<td>0.837</td>
<td>0.735</td>
<td>0.590</td>
<td>0.771</td>
</tr>
<tr>
<td>3.0</td>
<td>0.951</td>
<td>0.870</td>
<td>0.743</td>
<td>0.556</td>
<td>0.725</td>
</tr>
<tr>
<td>3.5</td>
<td>0.976</td>
<td>0.915</td>
<td>0.783</td>
<td>0.556</td>
<td>0.713</td>
</tr>
<tr>
<td>4.0</td>
<td>0.990</td>
<td>0.954</td>
<td>0.838</td>
<td>0.523</td>
<td>0.726</td>
</tr>
<tr>
<td>4.5</td>
<td>0.997</td>
<td>0.979</td>
<td>0.892</td>
<td>0.628</td>
<td>0.754</td>
</tr>
<tr>
<td>5.0</td>
<td>0.999</td>
<td>0.992</td>
<td>0.938</td>
<td>0.685</td>
<td>0.792</td>
</tr>
<tr>
<td>5.5</td>
<td>1.000</td>
<td>0.998</td>
<td>0.968</td>
<td>0.747</td>
<td>0.834</td>
</tr>
<tr>
<td>6.0</td>
<td>1.000</td>
<td>0.999</td>
<td>0.985</td>
<td>0.809</td>
<td>0.875</td>
</tr>
<tr>
<td>7.0</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>0.907</td>
<td>0.940</td>
</tr>
</tbody>
</table>
Figure 5. Efficiency of \( M \) for \( r = 1.00 \):

- \( 0.10 = q(\xi) \)
- \( 0.25 = q(\xi) \)
- \( 0.50 = q(\xi) \)
- \( 0.00 = q(\xi) \)

Efficiency relative to \( \bar{x} \).
3. Effects of varying \( a \) when \( b \) is held constant

In order to observe the effects of varying the parameter \( a \), it was necessary to select a value for \( b \) and then to assign various values to \( a \). The curves of Figure 6 are the efficiencies of \( W_{ab}(t) \), relative to \( \bar{X}_1 \), as functions of \( |\gamma| \) for \( b = .10 \) and for \( a = 1.00, .65, .42 \). These curves show that decreasing \( a \) when \( b \) is held constant; (1) decreases the efficiency at \( |\gamma| = 0 \), (2) decreases the maximum possible loss of efficiency, (3) shortens the range of \( |\gamma| \) for which a large loss of efficiency occurs, and (4) increases the effective difference.

4. Comparison of the efficiencies of \( W_o(t), W_{ab}(t), \) and \( SP(t) \)

The relative efficiencies of \( W_o(t), \) of \( W_{ab}(t) \) with \( a = .42 \) and \( b = .10 \), and of \( SP(t) \) with \( t_a = 1.6 \) are plotted as functions of \( |\gamma| \) in Figure 7. The constants involved in the last two estimators were selected so that the efficiencies of all three would be very nearly equal for \( |\gamma| \) equal to zero. The graph reveals the following important facts: (1) the greatest effective difference is attained with \( W_{ab}(t) \) and the smallest with \( SP(t) \), (2) the maximum loss is smallest for \( W_o(t) \) and largest for \( SP(t) \), and (3) the range of \( |\gamma| \) for which large losses occur is shortest for \( SP(t) \).

To compare these estimators numerically on the basis of overall efficiency as defined in Section VI, the integral

\[
I = -\frac{\sigma^2}{2n} \int_{-\infty}^{\infty} t^2 \left[ 1 - \theta(t) \right]^2 dt,
\]
Figure 6. Efficiency of $W_{ab}(t)$ for $b = .10$.

1. $a = .42$
2. $a = .65$
3. $a = 1.00$
Figure 7. Efficiency of $\mu_2^q, \mu_2^q(1)$, and $\mu_2^q(3)$.
was evaluated for each of the three weighting functions with the following results:

\[ I_{SP} = -\frac{\sigma^2}{2n} \int_{-1.6}^{1.6} t^2 dt = -1.365 \frac{\sigma^2}{n} \]

\[ I_{W_0} = -\frac{\sigma^2}{2n} \int_{-\infty}^{\infty} \frac{t^2 dt}{(1+t^2)^2} = -0.785 \frac{\sigma^2}{n} \]

\[ I_{W_{ab}} = -\frac{\sigma^2}{2n} \int_{-\infty}^{\infty} 1.1764t^2 e^{-2t^2} dt = -0.875 \frac{\sigma^2}{n} \]

Of the three estimators under consideration, \( W_0(t) \) has the greatest overall efficiency in the sense that the area between \( D_{NP}(\gamma) \) and \( D_W(\gamma) \) is largest for \( W_0(t) \); that is, the integral,

\[ \int_{-\infty}^{\infty} \left[ D_{NP}^2(\gamma) - D_W^2(\gamma) \right] d\gamma \]

is largest for \( W_0(t) \).
VIII. TRANSFORMED CORRELATION COEFFICIENTS

A. The Transformation

As the distribution of the sample correlation coefficient, $r$, based on a random sample of size $n$ from a bivariate normal distribution, is cumbersome to deal with in practical work, the correlation coefficient is frequently transformed in such a manner that the variance of the new variable is independent of the population correlation coefficient, $p$. This leads to the variable

$$
    z = \frac{1}{2} \ln \frac{1+r}{1-r}, \quad -1 \leq r \leq 1,
$$

which, even for small samples, is normally distributed with good approximation and which has mean

$$
    \mu = \frac{1}{2} \ln \frac{1+p}{1-p} + \frac{p}{2(n-1)},
$$

and variance

$$
    \sigma^2 = \frac{1}{n-3}.
$$

The transformation of $r$ may be used to test whether two sample correlation coefficients differ significantly. Let $r_1$ and $r_2$ be two independent sample correlation coefficients from samples of size $n_1$.
and \( n_2 \) respectively. The hypothesis to be tested is \( \rho_1 = \rho_2 = \rho \) or \( \mu_1 = \mu_2 = \mu \), so that, if the hypothesis is true,

\[
t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}
\]

is approximately normally distributed with mean zero and variance one.

B. The Pooling Procedure

If there is reason to believe that \( \rho_1 \) and \( \rho_2 \) are equal or nearly equal and if \( r_1 \) and \( r_2 \) are to be pooled to obtain an estimator for \( \rho_1 \) which, it is hoped, will have a smaller variance about \( \rho_1 \) than does the "never pool" statistic \( r_1 \), the use of the transformation (31) and the rule of procedure of Section V leads to the estimator

\[
Z_W(t) = \theta(t)z_1 + \left[ 1 - \theta(t) \right] \frac{(n_1 - 3)z_1 + (n_2 - 3)z_2}{n_1 + n_2 - 6}
\]

for the mean of the transformed coefficient \( z_1 \). In terms of \( r_1 \) and \( r_2 \), \( Z_W(t) \) is

\[
Z_W(t) = \frac{(n_2 - 3)\theta(t) + n_1 - 3}{Z(n_1 + n_2 - 6)} \ln \frac{1 + r_1}{1 - r_1} + \frac{(n_2 - 3) \left[ 1 - \theta(t) \right]}{Z(n_1 + n_2 - 6)} \ln \frac{1 + r_2}{1 - r_2},
\]

where

\[
t = \frac{z_1 - z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}
\]
is approximately normally distributed with mean

\[ \gamma = \frac{\mu_1 - \mu_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}} \]

and variance one. \( \theta(t) \) is a function of \( t \) only. The means are

\[ \mu_1 = \frac{1}{2} \ell n \left( \frac{1 + \rho_1}{1 - \rho_1} \right) + \frac{\rho_1}{2(n_1-1)} \]

and

\[ \mu_2 = \frac{1}{2} \ell n \left( \frac{1 + \rho_2}{1 - \rho_2} \right) + \frac{\rho_2}{2(n_2-1)} \]

The bias of \( Z_{W}(t) \), as an estimator of

\[ \bar{\gamma}_1 = \frac{1}{2} \ell n \left( \frac{1 + \rho_1}{1 - \rho_1} \right) \]

can be shown to be equal to

\[ B_{Z}(\gamma, \rho_1) = \frac{n_2 - 3}{(n_1-3)(n_1+n_2-6)} \left[ \gamma - \int_{-\infty}^{\infty} t\theta(t)N(t-\gamma)dt \right] - \frac{\rho_1}{2(n_1-1)} \]

The bias is composed of two parts. The first term is the bias which results from pooling when \( \rho_1 \neq \rho_2 \); the second term arises from the bias of \( z_1 \) as an estimator for \( \bar{\gamma}_1 \). In the pooling procedure no attempt to correct for the former will be made. In practical applications of pooling transformed correlation coefficients it is customary to attempt a correction for the bias in the pooled estimate for \( \rho_1 \) by solving for \( r \) in the equation
\[ \frac{1}{Z} \ln \frac{1 + r}{1 - r} = z - \frac{r_1}{2(n-1)}, \]

where \( r_1 \) is a first approximation for \( r \) obtained from the solution of

\[ \frac{1}{Z} \ln \frac{1 + r}{1 - r} = z. \]

Following the procedure outlined above for correcting for bias, the pooled estimate for \( \rho_1 \) is obtained from \( Z_w(t) \) by solving for \( R_w(t) \) in the equation

\[ \frac{1}{Z} \ln \frac{1 + R_w(t)}{1 - R_w(t)} = Z_w(t) - \frac{R_1}{2(n_1-1)}, \]

where \( R_1 \) is a first approximation resulting from the solution of the equation,

\[ \frac{1}{Z} \ln \frac{1 + R_1}{1 - R_1} = Z_w(t). \]

C. Examples

Snedecor (12, p. 151) gives the correlations between gain in weight and amount of feed eaten for two lots of pigs. The data (Table 7.4 in Snedecor) are reproduced in part in Table 5.

Substituting the values from the table,

\[ t = \frac{1.333 - .633}{\sqrt{.611}} = \frac{.700}{.782} = .895. \]
Table 5. Two correlations of gain with feed eaten among swine

<table>
<thead>
<tr>
<th>Lot</th>
<th>Pigs in Lot</th>
<th>r</th>
<th>z</th>
<th>1/(n-3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>.870</td>
<td>1.333</td>
<td>.500</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>.560</td>
<td>.633</td>
<td>.111</td>
</tr>
</tbody>
</table>
Let 
\[ \theta(t) = \frac{t^2}{1 + t^2} = \frac{(0.895)^2}{1 + (0.895)^2} = 0.445, \]
then, from Equation (32),
\[ Z_W(t) = (0.445)(1.333) + (0.555) \frac{2(1.333) + 9(0.633)}{11}, \]
\[ Z_W(t) = 0.593 + 0.422 = 1.015, \]
and
\[ R_1 = \frac{e^{2.03} - 1}{e^{2.03} + 1} = \frac{6.614}{8.614} = 0.768. \]

Then, using the first approximation, \( R_W \) is found by solving the equation
\[ \frac{1}{Z} \ln \frac{1 + R_W}{1 - R_W} = 1.015 - \frac{0.768}{8} = 0.919. \]

The estimate for \( \rho_1 \) obtained by using the weighting procedure is
\[ R_W = \frac{e^{1.838} - 1}{e^{1.838} + 1} = \frac{5.236}{7.236} = 0.724. \]

If the "sometimes pool" procedure based on the test of the hypothesis that \( \mu_1 = \mu_2 \) is used, and if the critical region is taken to be \( |t| > 1.6 \), the hypothesis is accepted, since \( t = 0.895 \), and \( \theta(t) = 0 \). The estimate for \( \mu_1 \) is
\[ Z_{SP}(t) = \frac{(n_1 - 3)z_1 + (n_2 - 3)z_2}{n_1 + n_2 - 6} = 0.76. \]

The first approximation to \( R_{SP} \) is
The pooled estimate for $p_1$ is

$$R_{SP} = \frac{e^{1.36} - 1}{e^{1.36} + 1} = \frac{2.896}{4.896} = .592.$$ 

The difference between the two estimates for $p_1$ arrived at by using the weighting procedure and the preliminary testing procedure can be explained as follows. In the weighting procedure if $t$ is not zero full weight is never given to the "always pool" statistic. The weighting procedure implies that, though $t$ is not significant, the difference between $p_1$ and $p_2$ may be small but it is not necessarily zero. Even so, some of the information contained in $r_2$ may be used to aid in estimating $p_1$. In the preliminary testing procedure a non-significant $t$ leads to accepting the hypothesis that $p_1 = p_2 = \rho$ and the "always pool" estimator is used to estimate the common value, $\rho$. Thus, even though $r_1$ is based on less than half of the number of observations used to find $r_2$, sufficient weight is given to $r_1$ in the weighting procedure so that $R_W$ is closer to $r_1$ than is $R_{SP}$, which is based on a weighted mean of the $z$'s.

If the approximation is good the mean square deviations and biases of $Z_W(t)$ and $Z_{SP}(t)$ may be compared by referring to Equations (23) and (24) and Figures 1 and 4.
IX. DISCUSSION

A comparison of the efficiencies of the estimator $W_o(t)$ for $\mu_1$ based on the weighting procedure and the estimator $SP(t)$ for $\mu_1$ resulting from the preliminary testing procedure was presented in Table 1 and in Figure 1 for the case of pooling the means of samples from normal populations when the means have a known common variance. This comparison indicates that the generalized pooling procedure, in this case, is effective in reducing the size of the maximum disturbance which may occur when the "sometimes pool" procedure is used and there is no a priori information available concerning the size of the nuisance parameter, $\gamma$. For values of $|\gamma|$ greater than approximately 3.2, $W_o(t)$ is less efficient than $SP(t)$ but the overall efficiency of $W_o(t)$ was shown, Section VII, to be greater than that of $SP(t)$. Furthermore, the effective difference of $W_o(t)$ is appreciably greater than that of $SP(t)$. The equations derived in Sections IV and V for the mean square deviations of $W(t)$ in the general case of normally distributed estimators with different known variances and in the various cases of pooling means with different known variances show that in these cases the weighting procedure offers the same advantages over the preliminary testing procedure as outlined above for the special case.

The biases of the two procedures were compared in Table 2 and Figure 4. This comparison is not favorable to $W_o(t)$ for, although
the bias of $W_o(t)$ is slightly less than that of $SP(t)$ for small values of $|\gamma|$, the bias of $SP(t)$ approaches zero much more rapidly than the bias of $W_o(t)$. However, it is felt that the mean square deviation, or the efficiency calculated from the mean square deviation, is a better measure of the relative merits of the two estimators since it is the variance about the true expected value of the estimator plus the square of the bias.

In Section VI it was shown that there was no uniformly minimum variance weighting function and that no admissible weighting function except $\theta(t) = 1$ would yield an unbiased estimate for $\theta_1$. The choice of a weighting function might be based on one of the following criteria:

1. Select a weighting function which will provide a large effective difference.
2. Choose a weighting function which will provide a small maximum loss of efficiency.
3. Select a weighting function which will provide a large overall efficiency.
4. Select a weighting function such that there will be a large gain in efficiency if $\gamma$ is equal to zero.

The results obtained from the investigation of the effects of varying the parameters of the two-parameter family of weighting functions were presented in Tables 4 and 5 and Figures 5, 6, and 7. These results show that, (1) increasing the efficiency at $\gamma = 0$ increases the maximum possible loss of efficiency, and (2) increasing the effective difference results in a loss of overall efficiency. The
statement (2) can be verified by considering the following examples:

(1) If \( a = 1.00 \) and \( b = .50 \) the effective difference is, from Figure 5, about 1.1. The overall efficiency, as defined in Section VI, is 
\(- .443 \sigma^2 / n.\)

(2) If \( a = .6 \) and \( b = .25 \) the effective difference is approximately 1.4. The overall efficiency is 
\(- .451 \sigma^2 / n.\)

(3) If \( a = .42 \) and \( b = .10 \) the effective difference is 1.8. The overall efficiency is 
\(- .875 \sigma^2 / n.\)

(4) If \( \phi(t) = \phi_o(t) \) the effective difference is approximately 1.45. The overall efficiency is 
\(- .785 \sigma^2 / n.\)

The large difference in overall efficiency between the weighting functions of examples (2) and (4) above when the effective differences are nearly the same is accounted for by the fact that for \( \phi(t) \) the maximum possible loss of efficiency is about 10 percent less than that of \( \phi(t; .6, .25) \). The price paid for this reduction in the maximum loss is the decrease in overall efficiency. It would appear that some compromises are necessary when a weighting function is selected.

It is realized that only a beginning has been made on the applications and effects of the generalized pooling procedure and that the problems which were considered in this investigation belong to the simplest class of problems to which the procedure might be applied. The author feels, however, that the results which have been achieved here indicate that the procedure should be effective in controlling some of the disturbances which arise in other more complex applications of preliminary tests of significance. He feels that the advantages claimed
for the weighting procedure in this study warrant further investiga-
tions along two lines, (1) an investigation should be made into the
operative characteristics of the procedure when used in the other
problems for which the effects of a preliminary test have been
studied, and (2) a more rigorous examination of possible weighting
functions and rules for their selection should be considered.
X. SUMMARY

In this thesis a generalized procedure for pooling estimators was proposed in an attempt to reduce the magnitudes of the disturbances in statistical inferences which result from pooling when the decision to pool, or not to pool, is based on a preliminary test of significance. It was shown that in those estimation problems where it is applicable the generalized procedure includes the preliminary testing procedure as a special case. The investigation of the effectiveness of the generalized pooling procedure in reducing the disturbances resulting from the preliminary testing procedure was confined to comparing the biases and mean square deviations of the estimators based on the two procedures when both are applied in a particular type of problem.

The first problem considered was that of pooling the means of two independent random samples of the same size drawn from two normal populations which have a known common variance, the object being to obtain an estimator for the mean, \( \mu_1 \), of one of the populations. The pooled estimator may have a smaller variance about \( \mu_1 \) than the unbiased "never pool" statistic, \( \bar{X}_1 \), the mean of the sample from the population concerned. The generalized or weighting procedure resulted in the estimator \( W_0(t) \) when the weighting function which determines the weights to be assigned to each of the sample means was taken to be \( t^2/(1 + t^2) \). The estimator \( SP(t) \) was obtained from the preliminary
testing procedure. The mean square deviations and biases were
derived for both estimators and the efficiency of each relative to \( \overline{X}_1 \)
as a function of the nuisance parameter, \( \gamma \), was calculated for
various values of \( \gamma \). It was found that if the critical region of the
preliminary test is selected so that \( W_0(t) \) and \( SP(t) \) have the same
efficiency when \( \gamma \) is equal to zero the weighting procedure results
in a sizeable reduction in the maximum possible loss of efficiency.
It was found that the effective difference was larger for \( W_0(t) \) than
for \( SP(t) \). The effective difference was defined to be the largest
value of \(|\gamma|\) such that for all smaller \(|\gamma|\) a gain in efficiency results.
When the biases were compared it was found that for values of \(|\gamma|\)
less than approximately 1.8 the bias of \( W_0(t) \) is slightly less than
that of \( SP(t) \) but that for larger values of \(|\gamma|\) the bias is smaller for
\( SP(t) \).

The equations derived for the mean square deviation and bias were
generalized to include the pooling of independent normal estimators
with different known variance and, in particular, the mean square
deivation was derived for the case of pooling sample means when the
variances of the means are known but different because of unequal
population variances, different sample sizes, or both. It was found
that if the ratio of the variances of the means, \( V_1/V_2 \), is greater
than one, the efficiency will be greater than that obtained for the
special case, \( V_1/V_2 = 1 \), for all \(|\gamma|\) less than the effective differ-
ence, but for all \(|\gamma|\) greater than the effective difference the
efficiency will be less than when \( V_1/V_2 = 1 \).
An admissible class of weighting functions was defined and three criteria, unbiasedness, uniformly minimum variances, and overall efficiency, were considered for selecting a "best" weighting function. Theorems were stated and proved which showed that, (1) the only unbiased weighting function is \( \theta(t) = 1 \), (2) no uniformly minimum variance weighting function exists, and (3) the weighting function with the greatest overall efficiency is \( \theta(t) = 1 \) which yields the "never pool" estimator.

A two-parameter family of weighting functions was defined and the mean square deviation of the weighting procedure based on this family was derived. The efficiency relative to \( \bar{X}_1 \) was calculated for various values of \( \gamma \) and for various values of \( a \) and \( b \) for the case of equal means with known common variance. The results were compared with those obtained by using \( \theta_0(t) = t^2/(1 + t^2) \) for the weighting function. It was found that the effective difference can be increased by decreasing \( b \) and adjusting \( a \) to give the desired efficiency when \( \gamma \) is zero. This increase in effective difference, however, was found to be accompanied by an increase in the range of \( |\gamma| \) for which large losses of efficiency may occur. This suggests that in selecting a weighting function a compromise is required.

The pooling of two transformed correlation coefficients in order to estimate one of the population correlation coefficients was used to illustrate the application of the generalized procedure. The pooling procedure appropriate to this problem was outlined and was illustrated by means of a numerical example.
XI. LITERATURE CITED


XII. ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professor T. A. Bancroft for his constant encouragement and assistance during the preparation of this thesis, and to Professor J. A. Gurland for his helpful suggestions which led to briefer and more direct derivations for some of the results than the methods originally used.