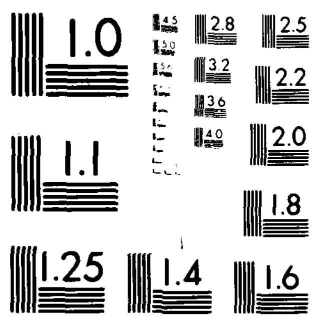
AD-A084 844 PRINCETON UNIV NJ DEPT OF STATISTICS F/G 12/1
USING BIWEIGHTS IN THE TWO-SAMPLE PROBLEM.(U)
NOV 79 K KAFADAR DAAG29-76-G-0298
UNCLASSIFIED TR-152-SER-2 ARO-14244.12-M NL

1 of 1
AD
A084844

END
DATE
FILMED
7 80
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA084844

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

12

LEVEL

1. REPORT NUMBER (19) 14244.12-M	2. GOVT ACCESSION NO. AD-A084844	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) USING BIWEIGHTS IN THE TWO-SAMPLE PROBLEM.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Karen/Kafadar		8. CONTRACT OR GRANT NUMBER(s) (15) DAAG29-76-G-0298
9. PERFORMING ORGANIZATION NAME AND ADDRESS Princeton University Princeton, NJ 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE (11) Nov 79
		13. NUMBER OF PAGES 37
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 28		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

(18) ARO

DTIC ELECTE MAY 28 1980 S D A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA (14) TR-152-SER-2

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

statistics
students T
confidence intervals
sampling

406873

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~We propose~~ It is proposed to ~~replacing~~ replace the usual Student's-t statistic which tests for equality of means of two distributions and ~~is used~~ to construct a confidence interval for the difference by a biweight-"t" statistic. The biweight-"t" is a ratio of the difference of the biweight estimates of location from the two samples to a standard error of this difference. Three different forms of the denominator are evaluated. Monte Carlo simulations reveal that resulting confidence intervals are highly efficient in moderate sample sizes, and that nominal levels are nearly attained, even when considering extreme percentage points.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DDC FILE COPY

USING BIWEIGHTS IN
THE TWO-SAMPLE PROBLEM

by

Karen Kafadar

Princeton University

Technical Report No. 152, Series 2
Department of Statistics
Princeton University
November 1979

This research was supported in part by a contract with the U. S. Army Research Office, No. DAAG29-76-0293, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

ABSTRACT

We propose replacing the usual Student's-t statistic which tests for equality of means of two distributions and is used to construct a confidence interval for the difference by a biweight-"t" statistic. The biweight-"t" is a ratio of the difference of the biweight estimates of location from the two samples to a standard error of this difference. Three different forms of the denominator are evaluated. Monte Carlo simulations reveal that resulting confidence intervals are highly efficient in moderate sample sizes, and that nominal levels are nearly attained, even when considering extreme percentage points.

ACKNOWLEDGEMENT

This report is based on sections in the author's Ph.D. dissertation, "Robust Confidence Intervals for the One- and Two- Sample Problems," Princeton University (1979). The author gratefully acknowledges Professors John W. Tukey and Peter Bloomfield for much helpful advice during its preparation and for numerous comments on early drafts of this report.

INTRODUCTION

The use of Student's-t in constructing confidence intervals for the difference in location of two populations is a common procedure. It is well known that this procedure is optimal when the underlying populations follow Gaussian distributions with the same variance. When the distributions are in fact even slightly stretched-tailed, however, studies show that, while the Student's-t interval nearly maintains its validity under the null hypothesis ([11], [18]), the power may be substantially reduced ([18]). In order to achieve "robustness of efficiency" in addition to "robustness of validity" (as defined in [15]), this study proposes the use of biweights in a two-sample "t"-like statistic, which we shall call "biweight-'t'". The biweight has shown a great deal of promise in earlier studies ([3], [4], [5]). The two-sample problem raises the issues of combining information on scale and on variance of the numerator of biweight-"t". We shall attempt to judge when such borrowing may be justified.

This report is divided into four parts. Part A deals with the case of equal sample sizes. Part B considers unequal sample sizes, for which a distinction between borrowed and unborrowed denominators may be made. Part C examines the performance of biweight-"t" when the samples do not have common scale factors (the conventional "unequal variances" case). Finally, an example is given along with conclusions and strategies for the two-sample case in Part D.

PART A: EQUAL SAMPLE SIZES.

1. Form of two-sample biweight-"t" and concepts.

Let

$$X_{11},\ldots,X_{1n_1} \sim F_1((x-\mu_1)/\sigma_1)$$

and

$$X_{21},\ldots,X_{2n_2} \sim F_2((x-\mu_2)/\sigma_2)$$

denote our samples from our two distributional situations. Then the two-sample biweight-"t" takes the form

$$\text{"t"} = (T_1-T_2)/S$$

where the squared denominator estimates the variance of the numerator

$$\widehat{\text{var}}(T_1-T_2)=S_1^2+S_2^2$$

For a definition of the biweight and its associated variance, the reader is referred to [12]; we mention here only the computational methods. For a single sample, $x_1,\ldots,x_n$, the biweight estimate of location, T_{bi}, is defined as the solution to the equation

$$\sum_{i=1}^{n} \psi((x_i - T_{bi})/(cs)) = 0 , \qquad (1)$$

where

$$\psi(u) = \begin{cases} u(1-u^2)^2 = u \cdot w(u) , & |u| \le 1 \\ 0 & \text{else} \end{cases}$$

Here, s is an estimate of scale from the sample $x_1, \ldots, x_n$, and c is a multiple of the scale. (A choice of c recommended in [12] is that for which the denominator, $c \cdot s$ is between 4σ and 6σ in the Gaussian case. In this study we will choose c such that $c \cdot s$ is roughly 6σ for the Gaussian.)

We may rewrite (1) in terms of the "weight function", $w(u)$, where

$$w(u) = \psi(u)/u ,$$

whence

$$T_{bi} = \frac{\sum_{i=1}^{n} x_i w(u_i)}{\sum_{i=1}^{n} w(u_i)} , \quad u_i = \frac{x_i - T_{bi}}{c \cdot s} \qquad (2)$$

Equation (2) permits an iterative solution. We start the iteration with a robust estimate of location (in this study, the median of the sample). The location estimate at the kth iteration, $T_{bi}^{(k)}$, $k \ge 1$, is found by

$$T_{bi}^{(k)} = \frac{\sum_{i=1}^{n} x_i w((x_i - T_{bi}^{(k-1)})/(c \cdot s))}{\sum_{j=1}^{n} w((x_j - T_{bi}^{(k-1)})/(c \cdot s))} \qquad (3)$$

In determining an estimate of scale to use in (3), former studies (see, e.g., [5], [10]) suggest the median absolute deviation from the median (MAD):

$$s^{(0)} = \text{med} \mid x_i - T_{bi}^{(0)} \mid .$$

For reasons to become clear later, Lax [10] showed that a more efficient scale estimate may be that using the functional form

$$s_{bi} = n^{1/2} \cdot (c_g s^{(0)}) \cdot q_{411}(\{\tilde{u}_i\}) \qquad (4)$$

where

$$\tilde{u}_i = \frac{x_i - T^{(0)}}{c_g \cdot s^{(0)}}$$

and

$$q_{411}(\{\tilde{u}_i\}) = \frac{\sum_{i=1}^{n} \psi^2(\tilde{u}_i)}{[\sum_{i=1}^{n} \psi'(\tilde{u}_i)][\max(1, -1 + \sum_{i=1}^{n} \psi'(\tilde{u}_i))]} \qquad (5)$$

Here, as before, $T^{(0)}$ is the median of the sample, $s^{(0)}$ is the MAD, and c_g is again chosen in order that $c_g \cdot s^{(0)}$ is approximately the desired multiple of σ in the Gaussian case. (Since $s^{(0)} \doteq (2/3)\sigma$ for a Gaussian sample, we choose $c_g = 9$ for this calculation.)

Finally, the denominator of our "t"$_{bi}$ statistic involves S_{bi}, where S_{bi}^2 estimates the variance of T_{bi}.

Huber ([6]) derives the theoretical asymptotic variance of T_{bi}, from which we may obtain a finite-sample approximation to it as

$$S_{bi}^2 = \hat{Var}(T_{bi}) = (c \cdot s_{bi}) \cdot q(\{u_i\}), \qquad (6)$$

where

$$u_i = \frac{x_i - T_{bi}}{c \cdot s_{bi}},$$

as in equation (4). Notice that, <u>in functional form</u>,

$$S_{bi}^2 = s_{bi}^2 / n,$$

just as

$$\hat{Var}(\bar{X}) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n(n-1)} = \frac{\text{classical sample } s^2}{n}$$

in the Gaussian case. However, s_{bi}^2 uses the median and the MAD in its computation, whereas S_{bi}^2 uses the more advanced location and scale estimates T_{bi} and s_{bi}. Notice also that $q_{411}(\{u_i\})$ as defined in (5) may be written

$$q_{411}(\{u_i\}) = \frac{\sum_{i=1}^{n} u_i^2 (1-u_i^2)^4}{[\sum_{i=1}^{n} (1-u_i^2)(1-5u_i^2)][\max(1, -1 + \sum_{i=1}^{n} (1-u_i^2)(1-5u_i^2))]}$$

The exponents of $(1-u_i^2)$, $(1-u_i^2)$, and $(1-5u_i^2)$, respectively, suggest the subscript and the name "411 widther" for s_{bi} (Equation (4)).

When we have two samples, we compute T_{bi} and S_{bi} for each sample. If we denote these by $T_{j,bi}$ and $S_{j,bi}$ $(j=1,2)$, our two-sample biweight-"t" statistic then takes the form

$$"t"_{bi} = \frac{T_{1,bi} - T_{2,bi}}{\sqrt{S^2_{1,bi} + S^2_{2,bi}}} . \qquad (7)$$

Performance of biweight-"t" will be evaluated on three different distributions:

- o Gaussian
- o One-Wild (4 observations from $N(0,1)$; 1 unidentified observation from $N(0,100)$)
- o Slash ($N(0,1)$ deviate/independent $U[0,1]$ deviate) .

These three situations are likely to cover a reasonably broad range of stretched-tailed behavior.

Robustness of efficiency may be evaluated in several ways. In this study, the success of biweight-"t" will be measured primarily in terms of "efficiency" of the expected confidence interval length (ECIL), i.e.,

$$\mathrm{eff}(\alpha) = \left[\frac{\mathrm{ECIL}_{min}(\alpha)}{\mathrm{ECIL}_{actual}(\alpha)}\right]^2$$

where $\mathrm{ECIL}_{actual}(\alpha)$ was defined by Gross ([5]) as

$$\mathrm{ECIL}_{actual}(\alpha) = 2 \cdot \alpha\text{%-point} \cdot \mathrm{ave}(\text{denominator of "t"}),$$

and $\mathrm{ECIL}_{min}(\alpha)$ is the shortest confidence interval we could expect for the given situation at hand. For the Gaussian, these are, of course, Student's-t intervals; an approximation (cf. [7]) is used for $\mathrm{ECIL}_{min}(\alpha)$ in the *One-Wild* and *Slash* situations.

Furthermore, for practical ease of use, we wish to approximate the distribution of biweight-"t" to a Student's-t with some chosen

number of degrees of freedom. Hence, we shall make the correspondence

(critical point , α) ---> degrees of freedom

The critical points of the distribution were all computed via a Monte Carlo swindle, the details of which may be found in [8]. There were either 640 or 1000 samples in the simulation for each of the three sampling situations.

2. Asymptotic Normality.

That "t" of (7) has an asymptotically normal distribution is clear by the following argument: for each population,

$$n^{1/2}(T_i-\mu_i) \xrightarrow{D} N(0, E\psi_i^2/(E\psi_i')^2),$$

where the subscript of ψ denotes the distribution; e.g.,

$$E\psi_i^2 = \int \psi^2((x-T_i)/(cs_i))\,dF_i(x) . \qquad (8)$$

Hence,

$$n^{1/2}((T_1-T_2)-(\mu_1-\mu_2)) \xrightarrow{D} N\left(0, \frac{E\psi_1^{\,2}}{(E\psi_1')^2} + \frac{E\psi_2^{\,2}}{(E\psi_2')^2}\right).$$

Since (c.f. [2])

$$n \cdot s_i^2 \xrightarrow{D} E\psi_i^2/(E\psi_i')^2$$

we have by Slutsky's theorem that

$$\frac{(T_1-T_2)-(\mu_1-\mu_2)}{(S_1^2+S_2^2)^{1/2}} \xrightarrow{D} N(0,1)$$

3. Borrowing Scales.

Since each of the biweights in the numerator and each of the variance estimates in the denominator of "t" requires an estimate of scale, we may consider a pooled estimate if we believe that both samples have common scale. As shown in [7], such a pooled estimate in the one-sample "t" can substantially reduce the variability in our results. In our simulation, all samples have been scaled to have unit width (the MLE for σ in the Slash situaton was chosen to be that of σ for the Gaussian which defines the density), so we expect favorable results from using all information (both samples) to estimate a common scale.

In Exhibit 1 we present the results of two-sample "t" when both samples have the same underlying distribution. Both M-estimates in the numerator have been scaled by s_{bor}, where

$$s_{bor}^2 = \frac{(2n)\sum_{j=1}^{2}\sum_{i=1}^{n}(x_{ij}-T_j)^2(1-u_{ij})^4}{\left(\sum_{j=1}^{2}\sum_{i=1}^{n}\psi'(u_{ij})\right)\left(-1+\sum_{j=1}^{2}\sum_{i=1}^{n}\psi'(u_{ij})\right)} \qquad (9)$$

$$u_{ij} = (x_{ij}-T_j)/(9\cdot s_j^{(0)})$$

$$s_j^{(0)} = \operatorname*{med}_{i,j}\left|x_{ij}-T_j^{(0)}\right| , \qquad T_j^{(0)} = \operatorname*{med}_{1\le i\le n} x_{ij} .$$

The subscript refers to a scale estimate which "borrows" width information from more than one sample.

4. Performance of two-sample "t" when $n_1 = n_2 \geq 10$.

Exhibit 1 reveals extremely high performance for $n_i \geq 10$. In particular, the resulting confidence intervals are trivially less efficient than if we knew the true underlying distribution (90% or higher). Furthermore, we are entitled to the full degrees of freedom in our approximation to a Student's-t distribution, across a broad range of α-levels.

To be conservative, we might wish to approximate "t"$_{bi}$ by a Student's-t on nine-tenths of the nominal degrees of freedom. As a check on the robustness of validity of this procedure, we plot the actual error rate in using these critical points (denoted by $t_{.9ndf}(\alpha)$); that is, we plot in Figures A and B

$$r_2(\alpha) = 100\cdot(\text{"t"}_{bi} > t_{.9ndf}(\alpha))/\alpha \quad \text{vs} \quad -\log_{10}(\alpha) .$$

These graphs (n=10 and n=20) reveal that the actual error rate is only slightly smaller than the nominal for commonly used α-levels of .05, .025, .01. As we go further into the tails, however, the actual error rates may be as low as 30% of the nominal (even lower for Slash, n=10). While the robustness of classical procedures for extreme α-levels has not been investigated, a comparison with the values in [1] indicates that this procedure is highly robust of validity at α = .05; presumably this robustness extends to the extreme α-levels as well.

5. Different distributions: the need to unborrow.

All three distributions in this study are derived from the Gaussian with unit variance. This fact, however, does not imply that a pooled scale is appropriate, as Exhibit 2 reveals. When our two samples do not both have the same underlying distributional shape, ECIL efficency is still high, but we are down to only 21 degrees of freedom (=.75(ndf)) when both n = 20 and a Gaussian or a One-Wild shape is paired with a Slash. This is not surprising, for Equation (9) is designed to estimate a common scale. A comparison of the distributions based on a different characteristic of width, such as a pseudo-variance quantity, shows that the Slash is considerably wider than either the Gaussian or the One-Wild (c.f. [13]). Hence, we must consider the possibility that a pooled estimate of scale may not be entirely appropriate.

Recall that our scale estimate (9) borrows from both samples and is used in four places in our "t"-statistic. In general, of course, we shall not know when we are entitled to borrow. More importantly, this pooled scale now violates the independence assumption in the numerator. It is true that the asymptotic normality as given in Section 2 depends only upon the consistency (not the dependence) of the scale estimates in the numerator $(T_1 - T_2)$. However, we shall be applying this result to relatively small sample sizes. While the dependence between numerator and denominator in the one-sample problem did not affect the efficiency of a biweight-"t" (c.f.[7]), it is not clear how the increased dependence in the numerator of (7) will alter the distribution on finite sample sizes.

To illustrate the effect of eliminating this dependence between the variables in the numerator, Exhibit 3 provides the analogous results to those in Exhibit 2, but now the biweights and variance estimates in the denominator are based on individual-sample estimates of scale. Curiously, despite the incompatibility of scales in the Gaussian-Slash and Gaussian-One Wild pairs, biweight-"t" with pooled scales gives higher ECIL efficiency and, when $n_1 = n_2 = 10$, more degrees of freedom. Biweight-"t" with unpooled scales gives between 5 and 13 more degrees of freedom when $n_1 = n_2 = 20$. Examining the results of Exhibits 2 and 3 together, we could be fairly confident in a comparison of two-sample "t" to a Student's-t on 0.9x(nominal degrees of freedom), if we knew when it is preferable to borrow! In the absence of this information, however, we notice that there is only a small amount to be gained in those cases where borrowing is helpful. Hence, using borrowed scales and a matching to 0.9x(nominal degrees of freedom) will perform adequately, for all but the extreme α-levels. To be absolutely confident down to the .001%-point when using unborrowed scales, we may wish to use 0.75x(ndf).

6. The case when $n_1 = n_2 = 5$: Results on slices.

Results in [9] indicate that the sporadic presence of low- and medium-weight samples of only five observations can occasionally lead to misleading estimates of scale, thereby affecting the distribution of "t"$_{bi}$. We recall from [9] the definition of <u>slice</u> for samples of size five: for W = sum of the biweight weights, a slice is defined by:

(a) a distributional situation;

(b) a sample size;

(c) a range of values for W = sum of the (biweight) weights.

For n=5, three ranges were determined by the cut-off values 3.3 and 4.3, yielding "low-weight", "medium-weight", and "high-weight" slices.

A similar phenomenon occurs here, as Exhibit 4 reveals. Based on the probabilities given in [9], we present in Exhibit 5 the estimated frequency of possible pairs of slices. Clearly, the most frequent combination in all cases is a high-weight sample from each population. In cases where both samples are low or medium weight, it is not clear that borrowing would necessarily result in a payoff. Even if the samples have the same underlying distributional shape, and therefore are commensurate on a width scale, we know that s_{bi} may yield very different estimates of scale for low-, medium-, or high-weight samples.

Since the most common Gaussian slice is high-weight (occurring 94% of the time), Exhibit 6 shows the values of matching degrees of freedom and ECIL efficiencies on high-weight slices from the

same underlying distribution, and when we have one sample each from the high-weight Gaussian and the Slash. In general, we may conclude that:

(i) borrowing does pay off if both samples are high-weight;

(ii) with the most likely case of two high-weight samples, we can approximate the distribution of "t"$_{bi}$ by a Student's-t on the nominal degrees of freedom;

(iii) when only one of the samples is high-weight, the gain in borrowing is not worth the potentially large loss if the distributions are not the same;

(iv) different distributions indicate that we are better off not borrowing;

(v) low- and medium-weight samples are likely to need some adjustments in the denominator, such as those recommended in [9].

Even conditionally, the case for n=5 is hardly an asymptotic situation. That we can do so well when both samples are high-weight (by far the most likely situation, occurring more than 95% of the time) is encouraging. Further results on the performance of "t"$_{bi}$ with adjustments in the denominator for low- and medium-weight slices are needed before specific recommendations in these cases can be made. We tend to speculate, however, that they may perform satisfactorily in the two-sample statistic as well.

PART B: UNEQUAL SAMPLE SIZES.

This case is treated separately, because of the sample size dependence of the variance estimates in the denominator. As in [9], we shall again refer to conditional results using the sum of the weights when needed.

7. Analogous two-sample statistic; asymptotic normality

If we believe that our biweights in the numerator have the same variance, a common assumption in the usual two-sample approach, we may wish to pool our variance estimates in a "borrowed" (via mean squares) denominator:

$$\hat{\mathrm{Var}}(T_1-T_2) = S^2_{bor} = \frac{n_1(n_1-1)S_1^2+n_2(n_2-1)S_2^2}{n_1+n_2-2}\cdot\left(\frac{1}{n_1}+\frac{1}{n_2}\right) \qquad (10)$$

A borrowed-"t" then takes the form:

$$\text{"t"}_{bor} = ((T_1-T_2)-(\mu_1-\mu_2))/S_{bor} \qquad (11)$$

Notice that in Equations (10) and (11) we may, but need not, choose to scale the biweights and the variance estimates using a pooled estimate as in Equation (9).

The denominator in (11) weights the variances of the factors in the numerator according to the sample size. Such an approach would not be reasonable if $\mathrm{Var}(T_1) \neq \mathrm{Var}(T_2)$. In such cases, we consider separate estimates of the variance in an *unborrowed denominator* (c.f. Welch's approach [15] to the Behrens-Fisher problem):

$$\text{"t"}_{unbor} = ((T_1-T_2)-(\mu_1-\mu_2))/(S_1^2+S_2^2)^{1/2} \qquad (12)$$

since the variance of the numerator may also be estimated by

$$S^2_{unbor} = S_1^2 + S_2^2 \qquad (13)$$

This distinction did not of course arise in Part A, for then (10) and (11) reduce to the same formula.

That the forms of two-sample "t" as in (10) and (12) do indeed have asymptotic normal distributions can be seen as follows. Following the lines of the argument in Section 2, we know that

$$U_i = \frac{n_i^{1/2}(T_i-\mu_i)}{\sqrt{E\psi_i^2/(E\psi_i')^2}} \xrightarrow{D} N(0,1), \quad i=1,2 \qquad (14)$$

where the notation for the expectations is defined in (9). Furthermore, if $F_1=F_2$, then the denominators in (14) are the same for both samples, so

$$\hat{\sigma}_v^2(n_1,n_2) = \frac{n_1(n_1-1)S_1^2 + n_2(n_2-1)S_2^2}{n_1+n_2-2} \xrightarrow{P} \frac{E\psi^2}{(E\psi')^2}$$

Hence, we have that "t"$_{bor}$ may be written

$$\text{"t"}_{bor} = \frac{n_1^{-1/2}U_1-n_2^{-1/2}U_2}{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)^{1/2}}\cdot\frac{\sqrt{E\psi^2/(E\psi')^2}}{\hat{\sigma}_v(n_1,n_2)}$$

$$= \left[\frac{U_1}{\sqrt{1+(n_1/n_2)}} - \frac{U_2}{\sqrt{1+(n_2/n_1)}}\right]\cdot\frac{\sqrt{E\psi^2/E(\psi')^2}}{\hat{\sigma}_v(n_1,n_2)}$$

If $n_1 \to \infty$ and $n_2 \to \infty$ in such a way that $n_1/n_2 \to 1$, then

$$\frac{1}{\sqrt{1+(n_2/n_1)}} \to \frac{1}{\sqrt{2}}, \quad \frac{1}{\sqrt{1+(n_1/n_2)}} \to \frac{1}{\sqrt{2}},$$

Hence, using Slutsky's theorem in conjunction with the convergence in distribution in (14), we conclude that "t"$_{bor}$ has an asymptotically normal distribution. The result on "t"$_{unbor}$ is similar.

8. Borrowing versus unborrowing: scales and denominators.

When we no longer have equal sample sizes, we might be cautious and prefer not to borrow estimates of either scale or biweight "variance". We know that such a cautious procedure may be quite wasteful of valuable information, especially when one sample has only five observations. On the other hand, biweight variances need not be the same for all distributions, and unwarranted borrowing in such cases may prove misleading. We need to know, then, under what circumstances we ought to borrow or ought not to borrow.

For comparison purposes we shall limit our attention to the efficiency of biweight-"t" at $\alpha = .001$ as representative of the behavior of "t"$_{bi}$ over the range $.00001 \leq \alpha \leq .05$. In Exhibit 7 we present these results, where the denominator of "t" is:

A) S_{bor}, borrowed scales: "complete borrowing";

B) S_{bor}, unborrowed scales: "incomplete borrowing";

C) S_{unbor}, unborrowed scales: "complete unborrowing".

When the distributions are the same, there is nearly always advantage to complete borrowing, as seen most dramatically when

both underlying densities are Gaussian. In these cases, we may again approximate the distribution of "t"$_{bi}$ by a Student's-t with the nominal degrees of freedom. When the distributions are not equal, a conservative matching would be 0.9x(nominal df). When one of the distributions is Slash, incomplete borrowing appears more successful.

Finally, we remark that there are some cases for which "t"$_{bi}$ in any of the three forms appears totally unsuccessful (e.g., Slash, n = 5, with anything else). This is primarily due to the fact that the biweight-"t" interval, like any robust confidence procedure, cannot provide the same high efficiency for low- and medium-weight samples as for the high-weight samples. When the smaller sample is restricted to be from the high-weight slice, efficiencies on the biweight-"t" intervals are slightly higher than those in Exhibit 7(B). A solution may well depend on an appropriate use of the weight distribution in these small samples.

PART C: UNEQUAL WIDTHS.

9. Unborrowed denominators.

When our samples have different scales, it is clear that a Welch-like unborrowed denominator of the form (12) is the safest approach. To evaluate the performance of biweight-"t" in the presence of slightly unequal widths, we multiply the observations of one of the distributions by either $\sqrt{2}$ or 2, yielding "variance" ratios between 2 and 4. A slight to moderate difference in scales was chosen to provide some indication of the effect on the results.

In Exhibit 8, we show some trials of "t"$_{unbor}$ either when

$$F_1 \neq F_2, \quad n_1 = n_2$$

or when

$$F_1 = F_2, \quad n_1 \neq n_2$$

(As in Exhibit 7, only the results for $\alpha = .001$ are shown.) Notice that our previous matching of the distribution to a Student's-t on 0.75x(nominal degrees of freedom) for unpooled scales would be conservative. This result is similar to the conservative nature of Welch's unborrowed t-statistic (e.g., as shown in [11],[16]). Approximating the distribution of "t"$_{unbor}$ by a Student's-t on 0.9x(ndf) instead, we see from Exhibit 8 that the nominal levels are nearly attained at the extreme percent points, but the actual levels are often less than half the nominal when the difference in scales is slight. Nonetheless, replacing means and sample variances by their biweight counterparts provides us with a practical procedure in the construction of confidence intervals that will not be grossly exaggerated by outlying values.

As a final comment to the interval problem on samples of varying widths, we mention the concept of transformation, a familiar data analytic tool in such situations. When comparing several batches of data, Tukey ([14], chapter 3,4) draws attention to the importance of choosing a form of expression for the data for which the amounts of spread are roughly the same across batches. Such re-expression may be useful in dealing with the "unequal variances" problem of this section. For example, Anscombe's ([1]) variance stabilizing transformations of Poisson data have been shown to produce more similarity in spread. The results of biweight-"t" discussed in Parts A and B (perhaps even the completely borrowed "t") may then be applied successfully to such re-expressed data.

PART D. AN APPLICATION AND CONCLUSIONS.

10. Comparing warp breaks: an example of scale-borrowing.

To gain some insight in deciding how and when we ought to borrow scales when constructing confidence intervals, we examine the data on the number of warp breaks from six different types of warp as given in Exhibit 9 of Chapter 3 in [14]. After some calculations on the raw counts, it appears that a transformation via square roots offers a more appealing view of the data. In Exhibit 9 we illustrate the square roots of the data by a comparison box-and-whiskers plot, and compute a biweight and scale estimate for each.

While the first and sixth batches appear to have different widths, batches 2,3,4, and 5 might reasonably be considered to have the same width. We therefore compute a borrowed scale for the thirty-six observations in these four groups.

If we wish to borrow denominators only, an estimate of $\hat{\sigma}$ would be

$$\hat{\sigma}=\sqrt{((.9105)^2+(1.0563)^2+(.9986)^2+(.9165)^2)/4} = 0.9783$$

on .9(32) = 28 d.f., and thus

$$s^2_{pool}=\hat{Var}(T_2)=\hat{Var}(T_3)=\hat{Var}(T_3)=\hat{Var}(T_4)=0.9783/\sqrt{9}=0.3234$$

For batches 1 and 6,

$$\hat{Var}(T_1) = 1.328/\sqrt{9} = 0.443$$

$$\hat{Var}(T_6) = 0.515/\sqrt{9} = 0.172$$

each on 0.75(8)=6 d.f. If we are interested in all $\binom{6}{2} = 15$ pairwise differences, a 95% confidence interval for any one of them could be guaranteed by using the 5%/(2·15) ≈ .17%-point of a Student's-t distribution. Comparing batches 1 and 2, for example, yields

$$(T_1-T_2) \pm t_{(6+28)}(.001)\sqrt{s_1^2 + s^2_{pool}}$$

$$= (6.547-4.826 \pm 3.348 \cdot 0.548 = (-0.114,3.556)$$

Comparing batches 2 and 5 provides a confidence interval of the form

$$(T_2-T_5) \pm t_{(28)}(.001)\sqrt{s^2_{pool} + s^2_{pool}}$$

$$= (4.826-5.299) \pm 3.408 \cdot 0.457 = (-2.032,1.086)$$

A confidence interval for the difference in average values for batches 1 and 6 uses completely unborrowed denominator:

$$(T_1-T_6) \pm t_{(6+6)}(.001)\sqrt{s_1^2 + s_6^2}$$

$$= (6.547-4.283) \pm 3.930 \cdot 0.475 = (0.398,4.130)$$

We mentioned in Part B that pooling scales in the calculation of the biweight estimates may provide increased efficiency, but may also increase the possibility of misleading scale estimates if in fact such pooling is unjustified. With samples of only nine observations each, such pooling may be profitable with batches 2,3,4,and 5. A pooled scale computed as in Eqn. (9) is 0.9411. In this particular example, there is little change in the resulting confidence intervals.

11. Concluding comments for the two-sample case.

This study investigated the performance of a two-sample "t" statistic when classical sample means and variances are replaced by their biweight counterparts. The hope is that the resulting statistic would have a distribution which could be well approximated by one from the Student's-t family, from which valid, yet efficient, confidence interval statements concerning the difference in centers can be made.

While we can be satisfied to a large extent with the approximation of the distribution of "t"$_{bi}$ to a Student's-t on .9(nominal df), even in the extreme tails, and the corresponding increase in the degrees of freedom with borrowed denominators, we find that the scaling issue has further impact here. We can choose either to pool estimates of scale (a wise move if in fact we have common underlying situations), or use separate estimates (slightly safer when the situations are different). In the former case, we can construct biweight-"t" intervals using 0.9x(nominal df); in the latter we use 0.75x(nominal df). In either case, the

efficiency of the procedure (in terms of relative length of the interval) is upwards of 70%. The same applies when $n_1 = n_2$, if we weight the variance estimates proportional to their sample sizes ("borrowed" denominator). Small sample sizes (n = 5) pose a problem only when the underlying population is extremely heavy-tailed (e.g., Slash).

A few trials of unborrowed denominators were run in situations where the samples did not have common width. For the most part, the 0.75x(ndf) matching is quite conservative, and the approximation to Student's-t on 0.9x(nominal df) could be safely recommended for all but perhaps the most extreme percent points (.01% and beyond). When the underlying situations are the same, we have better than 60% triefficiency out to the .5% point. When the situations are different, the efficiency decreases with the increased difference in the distribution (in terms of the "heaviness" of the tails).

While further insight into the nature of the weight distribution may suggest refinements, present results indicate that we may feel confident in constructing two-sample biweight-"t" intervals using tabulated Student's-t percent points as outlined above. An investigation of the performance of biweight-"t" in the presence of asymmetry is being considered.

Exhibit 1

Biweight-"t" with pooled scales

$F_1 = F_2,\ n_1 = n_2$

tail area α	Gaussian crit pt	Gaussian d.f.	Gaussian eff.	One-Wild crit pt	One-Wild d.f.	One-Wild eff.	Slash crit pt	Slash d.f.	Slash eff.
				n = 20					
.05	1.663	71.1	97.36	1.662	91.0	94.97	1.677	47.8	69.39
.025	2.002	57.9	97.28	1.996	67.2	94.98	2.004	54.7	71.22
.01	2.403	50.5	97.16	2.397	54.9	94.98	2.392	58.6	73.45
.005	2.684	50.5	97.08	2.679	49.9	94.81	2.660	60.8	75.12
.001	3.279	43.9	96.52	3.290	43.3	94.34	3.228	62.2	78.61
.0005	3.533	43.1	96.88	3.543	41.5	94.05	3.460	60.7	79.76
.0001	4.08	41.9	96.86	4.111	38.7	93.28	3.984	55.8	81.61
.00005	4.305	41.7	96.88	4.350	37.9	92.97	4.207	53.7	82.00
.000025	4.528	41.5	96.94	4.586	37.3	92.68	4.428	51.7	82.10
.00001	4.813	41.4	97.08	4.894	36.8	91.41	4.720	49.5	82.08
				n = 10					
.05	1.692	33.3	93.74	1.693	32.7	86.60	1.700	28.4	70.86
.025	2.053	26.8	93.49	2.052	27.0	86.89	2.020	40.8	75.48
.01	2.450	23.4	93.09	2.823	23.3	87.09	2.402	51.1	82.27
.005	2.818	22.1	93.09	2.823	21.7	87.11	2.674	52.3	82.27
.001	3.537	20.7	92.99	3.571	19.3	86.72	3.277	46.5	92.74
.0005	3.840	20.4	93.03	3.898	18.6	86.34	3.527	44.1	89.46
.0001	4.546	19.9	93.39	4.692	17.3	84.87	4.096	40.1	97.94
.00005	4.853	19.8	93.39	5.054	16.9	83.96	4.341	33.6	99.82
.000025	5.164	19.7	93.54	5.432	16.5	92.93	4.589	37.2	100.74
.00001	5.581	19.6	93.77	5.955	16.0	81.49	4.923	35.4	101.51

Note: Standard errors of critical points for $n = 20$ range from .007 (α = .05) to .092 (α = .00001); for $n = 10$, the corresponding range is from .009 to .162.

Exhibit 2

Biweight-"t" with pooled scales

$F_1 \neq F_2,\ n_1 = n_2$

tail area	Gaussian & One-Wild			Gaussian & Slash			One-Wild & Slash		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
				n = 20					
.05	1.665	76.1	96.09	1.734	17.9	91.65	1.725	20.0	87.59
.025	2.001	58.9	95.96	2.090	19.4	93.25	2.072	22.4	89.65
.01	2.404	49.7	95.73	2.523	20.5	94.83	2.492	24.1	91.83
.005	2.688	45.9	95.52	2.833	20.9	95.63	2.792	24.6	93.07
.001	3.303	41.0	94.89	3.523	21.3	96.43	3.456	24.7	94.68
.0005	3.555	39.7	94.56	3.812	21.4	96.56	3.735	24.6	95.01
.0001	4.125	37.5	93.69	4.468	21.7	96.78	4.376	24.3	95.34
.00005	4.366	36.8	93.28	4.747	21.9	96.84	4.652	24.2	95.28
.000025	4.606	36.2	92.85	5.024	22.1	96.82	4.929	24.1	95.06
.00001	4.921	35.5	92.27	5.389	22.3	96.66	5.298	24.0	94.50
				n = 10					
.05	1.772	12.9	98.97	1.759	14.3	92.03	1.749	15.6	84.42
.025	2.171	12.4	97.09	2.125	15.5	94.64	2.105	17.5	87.45
.01	2.675	12.2	94.60	2.579	16.3	97.21	2.546	18.5	90.42
.005	3.047	12.2	92.97	2.912	16.4	98.47	2.870	18.5	91.94
.001	3.902	12.3	89.69	3.695	15.8	98.52	3.631	17.4	92.50
.0005	4.275	12.4	88.36	4.049	15.4	97.25	3.979	16.7	91.29
.0001	5.176	12.6	85.21	4.943	14.4	91.66	4.895	14.9	84.76
.00005	5.580	12.6	83.91	5.358	14.0	88.65	5.333	14.2	81.12
.000025	5.995	12.7	82.68	5.782	13.8	85.71	5.780	13.8	77.76
.00001	6.557	12.7	81.25	6.343	13.7	82.29	6.361	13.7	74.19

Note: **Standard errors** of critical points range from .008 (α = .05) to .150 (α = .00001) for n = 20; for n = 10, the corresponding range is from .014 to .240.

Exhibit 3

Biweight-"t" with unpooled scales

$F_1 \neq F_2$, $n_1 = n_2$

tail area	Gaussian & One-Wild			Gaussian & Slash			One-Wild & Slash		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
				n = 20					
.05	1.672	56.7	95.68	1.703	27.2	72.88	1.694	32.1	72.57
.025	2.010	49.2	95.55	2.041	30.4	74.91	2.028	36.1	74.76
.01	2.414	44.2	95.33	2.450	31.9	77.10	2.431	37.3	77.10
.005	2.699	41.9	95.12	2.737	32.4	78.52	2.714	37.5	78.66
.001	3.316	38.6	94.49	3.368	31.9	80.89	3.327	37.1	81.66
.0005	3.570	37.7	94.17	3.633	31.2	81.45	3.580	36.5	82.64
.0001	4.142	36.1	93.31	4.251	29.3	81.94	4.157	34.9	84.42
.00005	4.384	35.6	92.91	4.519	28.6	81.89	4.401	34.4	85.56
.000025	4.624	35.1	92.49	4.788	28.0	81.71	4.644	34.0	85.56
.00001	4.940	34.7	91.94	5.145	27.5	81.28	4.962	33.7	86.06
				n = 10					
.05	1.785	11.8	97.10	1.768	13.3	73.54	1.760	14.2	71.66
.025	2.186	11.7	95.22	2.122	15.8	76.66	2.111	16.9	74.77
.01	2.690	11.7	93.12	2.569	16.9	79.07	2.554	17.9	77.24
.005	3.061	11.9	91.70	2.901	16.9	80.11	2.883	17.8	78.29
.001	3.924	12.1	88.25	3.681	16.1	80.13	3.662	16.6	78.18
.0005	4.301	12.2	86.89	4.037	15.6	78.98	4.022	15.9	76.82
.0001	5.196	12.4	84.14	4.937	14.5	74.17	4.958	14.3	71.02
.00005	5.595	12.5	83.06	5.355	14.1	71.63	5.399	13.8	68.02
.000025	6.008	12.6	81.92	5.782	13.8	69.15	5.850	13.5	65.24
.00001	6.574	12.6	80.44	6.349	13.7	66.30	6.439	13.3	62.24

Note: Standard errors for critical points are of the same size as those in Exhibit 2.

Exhibit 4

Biweight-"t" with pooled scales

$n_1 = n_2 = 5$

$F_1 = F_2$

tail area	Gaussian			One-Wild			Slash		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
.05	1.849	8.4	91.11	1.769	13.2	60.53	1.790	11.4	68.08
.025	2.348	7.3	96.91	2.248	9.4	59.41	2.269	8.8	77.74
.01	3.100	6.3	78.60	2.973	7.2	56.03	3.110	6.2	100.52
.005	3.832	5.6	69.08	3.631	6.3	52.27	4.095	4.9	110.11
.001	7.267	4.0	34.54	6.483	4.5	32.33	7.927	3.7	113.12
.0005	10.046	3.6	22.67	9.528	3.8	19.60	12.035	3.2	91.32
.0001	16.658	3.6	13.49	16.326	3.7	12.11	28.573	2.9	47.03
.00005	19.325	3.7	12.22	18.532	3.8	12.03	46.825	2.5	3.05
.000025	21.873	3.8	11.61	20.476	3.9	12.59	56.648	2.6	2.30
.00001	25.061	3.9	11.37	22.755	4.1	14.01	66.471	2.7	1.83

$F_1 \neq F_2$

tail area	Gaussian & One-Wild			Gaussian & Slash			One-Wild & Slash		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
.05	2.213	3.5	54.74	2.163	3.8	62.74	1.975	5.5	50.75
.025	3.137	3.1	42.46	3.005	3.4	58.16	3.870	3.8	50.91
.01	5.645	2.4	21.07	4.816	2.8	48.72	3.870	3.8	50.91
.005	10.798	1.9	7.85	7.299	2.5	40.32	5.278	3.4	52.02
.001	38.305	1.8	1.17	25.437	2.0	9.06	11.178	2.9	31.65
.0005	52.171	1.9	0.80	46.698	1.9	3.30	15.201	2.9	21.03
.0001	84.351	2.0	0.52	96.043	2.0	0.99	28.161	2.9	7.76
.00005	98.208	2.0	0.43	117.291	2.0	0.73	34.149	2.9	5.81
.000025	112.065	2.3	0.46	138.538	2.0	0.57	40.138	3.0	4.62
.00001	130.382	2.3	0.46	166.625	2.4	0.45	48.053	3.0	3.67

Note: Standard errors for critical points range from .022 (α = .05) to 1.182 (α = .00001) when $F_1 = F_2$, and approximately twice this size when $F_1 \neq F_2$.

Exhibit 5

Frequency of pairs of slices when $n_1 = n_2 = 5$
(Expected number of cases per 1000)

$F_1 = F_2$

Gaussian,Gaussian

	L	M	H
L	0.6	0.7	23
M	0.7	0.8	27
H	23	27	899

One-Wild,One-Wild

	L	M	H
L	0.8	13	14
M	13	208	235
H	14	235	266

Slash,Slash

	L	M	H
L	2	10	34
M	10	45	157
H	34	157	552

$F_1 \neq F_2$

Gaussian (rows), One-Wild (columns)

Gaussian	L	M	H
L	0.7	11	12
M	0.8	13	14
H	27	432	489

Gaussian (rows), Slash (columns)

Gaussian	L	M	H
L	1	5	18
M	1	6	21
H	44	200	704

One Wild (rows), Slash (columns)

One Wild	L	M	H
L	1	6	21
M	21	96	339
H	24	109	383

L=Low-weight M=Medium weight H=High-weight

Exhibit 6A

Biweight-"t" on HIGH-WEIGHT slices

$F_1=F_2$, $n_1=n_2=5$

tail area	Gaussian			One-Wild			Slash		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
				pooled scales					
.025	2.225	10.1	95.93	1.810	∞	54.83	1.868	∞	58.02
.01	2.825	8.9	94.90	2.212	∞	60.53	2.285	∞	94.49
.005	3.312	8.4	93.75	2.529	∞	64.48	2.607	160.4	138.23
.001	4.625	7.5	90.06	3.328	36.8	73.40	3.448	25.3	378.56
.0005	5.266	7.4	88.59	3.703	26.3	77.65	3.909	18.3	558.21
.0001	6.812	7.4	87.50	4.610	18.7	90.86	5.576	10.5	385.74
.00005	7.461	7.6	88.29	4.999	17.6	98.98	6.542	9.1	312.00
.000025	8.089	7.8	89.80	5.379	17.0	109.08	7.505	8.6	259.78
.00001	8.883	8.0	92.78	5.868	16.7	126.02	8.682	8.3	215.33
				unpooled scales					
.05	1.850	8.3	96.16	1.519	∞	44.33	1.582	∞	64.13
.025	2.307	8.0	95.65	1.817	∞	49.12	1.900	∞	81.62
.01	2.926	7.7	94.71	2.214	∞	54.57	2.311	∞	134.34
.005	3.422	7.5	93.81	2.526	∞	58.37	2.631	92.7	197.33
.001	4.706	7.3	91.49	3.317	38.6	66.75	3.479	23.4	540.41
.0005	5.314	7.2	90.75	3.695	26.8	70.42	3.960	17.1	790.58
.0001	6.793	7.4	90.71	4.636	18.2	81.15	5.741	9.8	528.89
.00005	7.433	7.6	91.61	5.045	17.0	87.77	6.759	8.7	424.90
.000025	8.065	7.8	93.10	5.447	16.4	96.10	7.771	8.1	352.22
.00001	8.878	8.0	95.95	5.963	16.0	110.25	8.989	7.9	291.86

Exhibit 6B
Biweight-"t" on slices
$F_1 \neq F_2$, $n_1 = n_2 = 5$

tail area	Gaussian (high) & Slash (low)			Gaussian (high) & Slash (medium)			Gaussian (high) & Slash (high)		
α	crit pt	d.f.	eff.	crit pt	d.f.	eff.	crit pt	d.f.	eff.
				pooled scales					
.05	3.397	2.0	74.75	2.148	3.9	82.95	1.665	76.3	71.00
.025	4.799	1.8	67.01	2.770	4.0	89.27	2.073	22.2	81.96
.01	7.234	3.0	63.45	3.650	4.2	110.67	2.663	12.6	106.92
.005	9.487	2.1	70.14	4.369	4.4	146.83	3.218	9.4	139.12
.001	14.223	2.6	85.15	6.627	4.4	174.13	8.803	3.4	50.73
.0005	15.935	2.8	83.38	8.116	4.3	142.67	52.923	1.9	1.73
.0001	19.528	3.3	70.28	12.676	4.1	74.04	154.907	1.9	0.25
.00005	20.981	3.6	67.04	14.689	4.2	60.71	198.822	2.0	0.17
.000025	22.411	3.8	64.55	16.636	4.4	52.00	242.735	2.0	0.13
.00001	24.299	4.0	62.45	19.070	4.6	45.01	300.783	2.0	0.09
				unpooled scales					
.05	4.334	1.3	86.52	2.263	3.3	76.23	1.657	130.2	60.54
.025	5.805	1.6	86.28	2.867	3.7	84.95	2.010	48.8	73.59
.01	8.068	1.9	96.12	3.723	4.1	108.47	2.475	26.8	104.47
.005	9.989	2.0	119.21	4.411	4.3	146.83	2.833	21.0	151.57
.001	14.766	2.6	148.85	6.127	4.8	207.70	3.718	15.3	240.02
.0005	16.810	2.8	141.16	6.920	5.0	200.10	4.139	14.0	238.07
.0001	21.225	3.1	112.10	8.931	5.5	152.10	5.271	12.0	185.83
.00005	22.964	3.4	105.44	9.863	5.6	137.30	5.835	11.4	166.99
.000025	24.661	3.7	100.43	10.807	5.8	125.63	6.432	10.9	150.97
.00001	26.915	3.9	95.90	12.032	6.0	115.28	7.236	10.7	135.65

Note: Standard errors for critical points range from .010 (α = .05) to .470 (at α = .00001).

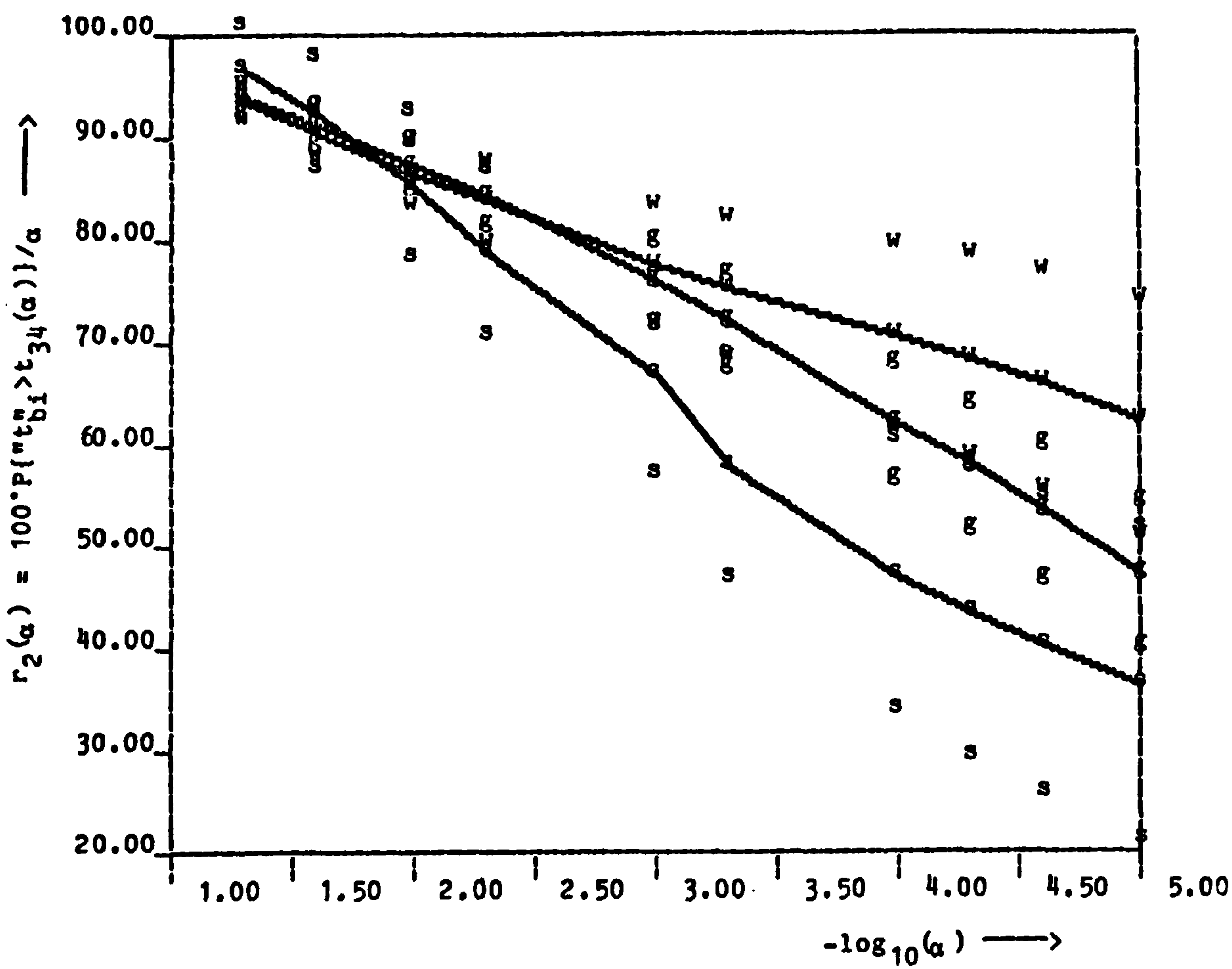
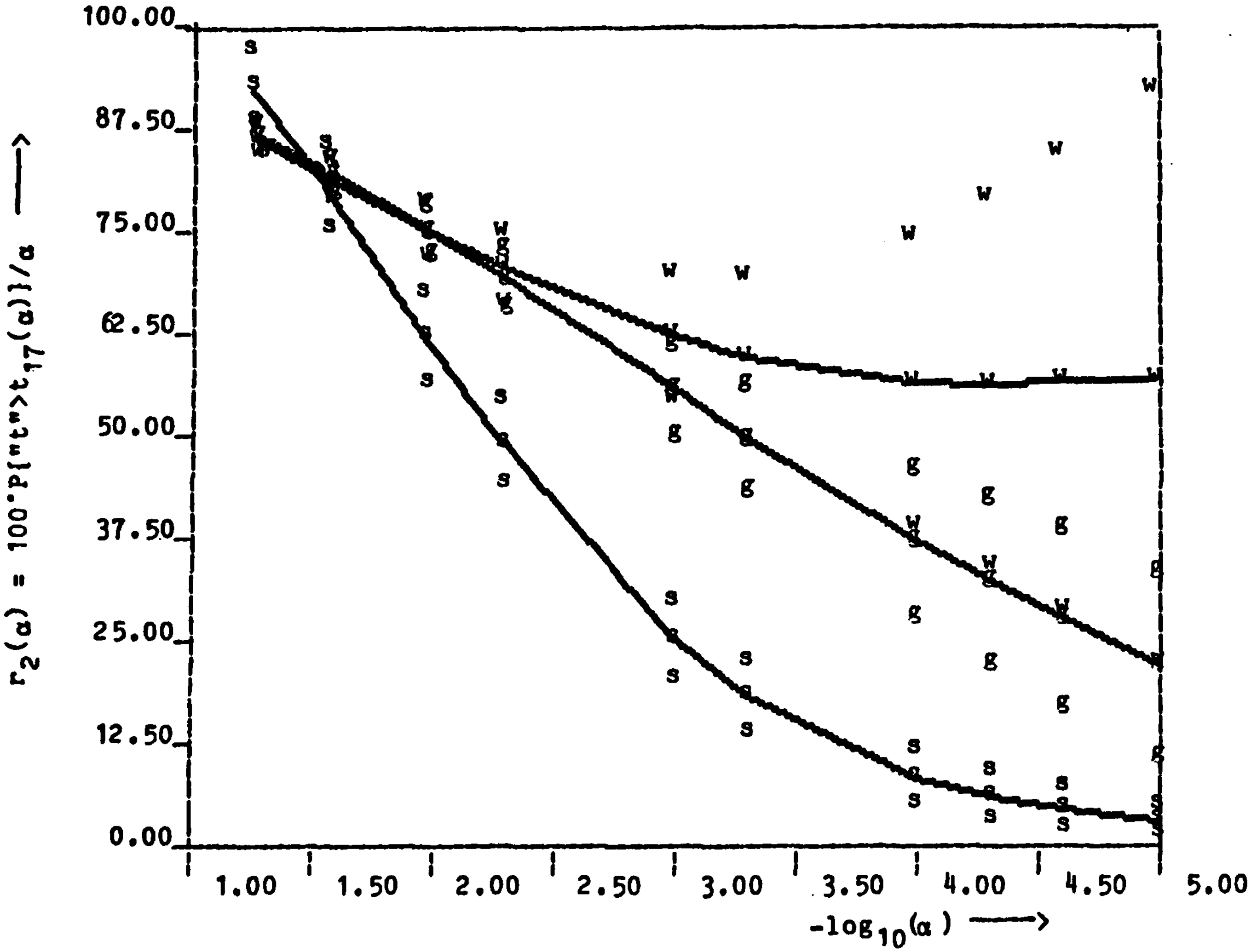


Figure A: Plot of $r_2(\alpha)$ vs $-\log(\alpha)$ for two-sample biweight-"t", $n_1 = n_2 = 20$, $F_1 = F_2$ (borrowed scales)
(g=Gaussian, w=One-Wild, s=Slash; shown with 1 std. dev. limits)



<u>Figure B</u>: Plot of $r_2(\alpha)$ vs $-\log(\alpha)$ for two-sample biweight-"t", $n_1=n_2=10$; $F_1=F_2$ (borrowed scales)
(g=Gaussian, w=One-Wild, s=Slash; shown with 1 std. dev. limits)

Exhibit 7

Matched degrees of freedom and ECIL efficiencies at α=.001
for two-sample biweight-"t": Unequal sample sizes(1)

F_1	n_1	F_2	n_2	complete borrowing df	complete borrowing eff	incomplete borrowing df	incomplete borrowing eff	complete unborrowing df	complete unborrowing eff	nominal df
A) $F_1 = F_2$ (2)										
G	10	G	20	30.0	96.13	25.5	94.45	17.4	85.11	28
W	10	W	20	29.4	91.95	24.3	88.18	16.8	76.08	28
S	10	S	20	27.9	79.99	21.7	73.75	37.1	80.45	28
G	5	G	10	14.5	95.71	10.6	85.78	6.4	58.24	13
W	5	W	10	11.8	75.93	9.9	65.40	6.3	37.72	13
S	5	S	10	13.7	384.56	10.9	330.41	5.5	211.06	13
G	5	G	20	24.1	95.19	14.8	83.92	4.7	32.06	23
W	5	W	20	14.2	78.95	12.5	70.84	4.7	19.98	23
S	5	S	20	10.0	427.34	10.1	419.14	5.5	211.06	23
B) $F_1 \neq F_2$										
G	10	W	20	40.1	96.35	30.2	92.82	17.7	84.23	28
G	10	S	20	∞	92.11	∞	82.20	36.3	92.64	28
W	10	G	20	23.3	91.55	21.1	88.84	16.2	75.46	28
W	10	S	20	∞	82.25	∞	78.43	40.6	88.57	28
S	10	G	20	8.4	99.68	8.6	82.52	14.4	65.23	28
S	10	W	20	9.0	99.90	9.1	83.94	16.9	68.84	28
G	5	W	10	13.9	87.75	10.7	78.16	6.6	59.28	13
G	5	S	10	59.0	77.11	53.2	64.57	11.6	75.53	13
W	5	G	10	12.4	84.69	9.5	67.34	6.0	35.29	13
W	5	S	10	19.2	50.38	28.1	53.78	9.8	52.83	13
S	5	G	10	1.9	14.70	4.4	273.40	3.6	102.15	13
S	5	W	10	1.9	11.54	4.6	267.82	3.7	105.87	13
G	5	W	20	27.6	92.41	15.9	80.67	4.7	32.10	23
G	5	S	20	∞	81.62	∞	72.50	12.8	74.29	23
W	5	G	20	15.2	86.53	11.6	71.42	4.6	18.78	23
W	5	S	20	∞	56.39	∞	56.30	11.8	48.96	23
S	5	G	20	2.0	38.22	4.6	489.44	3.1	73.52	23
S	5	W	20	2.0	44.16	4.6	467.52	3.1	74.78	23

(1) Standard errors for critical points from which degrees of freedom were matched and ECIL efficiencies were calculated fell in the range 0.028 to 0.331 for α = .001.

(2) F_j represents underlying distribution for sample j: G = Gaussian, W = One-Wild, S = Slash.

Exhibit 8

Matched degrees of freedom and ECIL efficiency for biweight-"t" at $\alpha = .001$: "σ_1" $\neq$ "σ_2"

F_1 [1]	n_1	F_2	n_2	$\left(\frac{\sigma_2}{\sigma_1}\right)^2$	matched d.f.	ECIL eff.	$\left(\frac{\text{actual } \alpha}{\text{nominal } \alpha}\right)$
A) $F_1 = F_2$							
G	10	G	20	2	∞ [2]	90.90	.341
G	10	G	20	4	∞	92.46	.038
W	10	W	20	2	∞	82.94	.379
W	10	W	20	4	∞	85.33	.045
S	10	S	20	2	∞	262.61	.255
S	10	S	20	4	∞	208.25	.073
G	20	G	10	2	∞	72.46	.469
G	20	G	10	4	∞	56.07	.286
W	20	W	10	2	∞	62.26	.593
W	20	W	10	4	∞	47.62	.474
S	20	S	10	2	∞	224.64	.177
S	20	S	10	4	∞	155.83	.030
B) $F_1 \neq F_2$							
G	20	W	20	2	∞	91.21	.127
G	20	W	20	4	∞	85.55	.007
G	10	W	10	2	∞	47.00	1.152
G	10	W	10	4	13.7	52.95	.533
G	20	S	20	2	∞	57.22	.238
G	20	S	20	4	∞	35.84	.038
W	10	S	10	2	∞	57.20	.307
W	10	S	10	4	∞	35.60	.038

Notes:

(1) F_j represents underlying distribution for sample j: G = Gaussian, W = One-Wild; S = Slash.

(2) Indicates that biweight-"t" distribution is shorter-tailed than Gaussian.

Exhibit 9
Summary of Tippet's warp breaks

A) The data.

26	18	36	27	42	20
30	21	21	14	26	21
54	29	24	29	19	24
25	17	18	19	16	17
70	12	10	29	39	13
52	18	43	31	28	15
51	35	28	41	21	15
26	30	15	20	39	16
67	36	26	44	29	28

B) Square-root transformation of the data

5.099	4.243	6.000	5.196	6.481	4.472
5.477	4.582	4.582	3.742	5.099	4.582
7.348	5.385	4.899	5.385	4.359	4.899
5.000	4.123	4.243	4.359	4.000	4.123
8.367	3.464	3.162	5.385	6.245	3.606
7.211	4.243	6.557	5.568	5.292	3.873
7.141	5.916	5.292	6.403	4.582	3.873
5.099	5.477	3.873	4.472	6.245	4.000
8.185	6.000	5.099	6.633	5.385	5.292

C) Comparison box-plots of transformed data.

```
8.367
             x
             |
             |
             |
7.066       -*-
            | |
            | |          x
            | |      x   |   x
            | |      |   |  -+-
5.764       | |  x   |   |  | |
            | |  |   |  -+- | |
            | | -+- -+- |*| |*|  x
            -+- | | | | | | | |  |
                | | |*| | | | |  |
4.463           |*| | | -+- -+- -+-
                -+- -+-  |   |  |*|
                 |   |   |   x  -+-
                 |   |   x       x
                 x   |
3.162                x
```

Exhibit 9 (con't.)

D) Summary statistics.

batch	n	$\hat{\mu}$	$\hat{\sigma}$	W=Σweights
1	9	6.547	1.328	8.531
2	9	4.822	0.856	8.526
3	9	4.853	0.976	8.507
4	9	5.241	0.886	8.515
5	9	5.299	0.849	8.524
6	9	4.283	0.515	8.508

REFERENCES

[1] Anscombe, F.J. (1948). The transformation of Poisson, binomial and negative binomial data. Biometrika 35, 246-254.

[2] Carroll, Raymond J. (1979). On almost-sure expansions for M-estimates. Ann. Statist. 6, No. 2, 314-318.

[3] Denby, L. and Larsen, W.A. (1977). Robust regression estimators compared via Monte Carlo. Communications in Statistics A6, 335-362.

[4] Denby, L. and Martin, R. Douglas (1979). Robust estimation of the first-order autoregressive parameter. J. Amer. Statist. Assoc. 74, No. 365, 140-146.

[5] Gross, A.M. (1976). Confidence interval robustness with long-tailed symmetric distributions. J. Amer. Statist. Assoc. 71, 409-417.

[6] Huber, Peter (1977). Robust Statistical Procedures. Philadelphia: Society for Industrial and Applied Math. (Regional Conf. Series in Applied Math., 27.)

[7] Kafadar, Karen (1979a). A biweight approach to the one-sample problem. Technical Report No. 151, Dept. of Statistics, Princeton University, Princeton, New Jersey.

[8] ____________ (1979b). A two-sample Monte Carlo swindle. Technical Report No. 153, Dept. of Statistics, Princeton University, Princeton, N. J.

[9] ____________ (1979c). A robust confidence interval for samples of five observations. Technical Report No. 154, Dept. of Statistics, Princeton University, Princeton, N. J.

[10] Lax, David (1975). An interim report of a Monte Carlo study of robust estimates of width. Technical Report No. 93, Dept. of Statistics, Princeton University, Princeton, N. J.

[11] Lee, Austin F.S. and D'Agostino, Ralph B. (1976). Levels of significance of some two-sample tests when observations are from compound normals. Communications in Statistics A5, No. 4, 325-342.

[12] Mosteller, F. and Tukey, J. W. (1972). Data Analysis and Regression: A second course in statistics. Addison-Wesley: Reading, Mass.

[13] Rogers, W.H. and Tukey, J. W. (1972). Understanding some long-tailed symmetrical distributions. Statistica Neerlandica 26, No. 3, 211-226.

[14] Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley: Reading, Mass.

[15] Tukey, J.W. and McLoughlin, Donald H. (1963). Less vulnerable confidence and significance procedures for location based on a single sample: Trimming/Windsorization. Sankhya, Series A, 25, 331-352.

[16] Welch, B.F. (1938). The significance of the difference between two means when the population variances are unequal. Biometrika 29, 350-362.

[17] Yuen, Karen K. (1974). The two-sample trimmed t for unequal population variances. Biometrika 61, Vol. 1, 165-169.

[18] Yuen, Karen K. and Dixon W.J. (1973). The approximate behavior and performance of the two-sample trimmed t. Biometrika 60, Vol. 2, 369-374.