AD-A088 071 PRINCETON UNIV NJ DEPT OF STATISTICS F/G 12/1

A ROBUST CONFIDENCE INTERVAL FOR SAMPLES OF FIVE OBSERVATIONS.(U)

NOV 79 K KAFADAR DAAG29-76-G-0298

UNCLASSIFIED TR-154-SER-2 ARO-14244.14-M NL

1 of 1

AD A088071

END

DATE FILMED

9-80

DTIC

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 14244.14-M	2. GOVT ACCESSION NO.: AD-A088071
	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): A ROBUST CONFIDENCE INTERVAL FOR SAMPLES OF FIVE OBSERVATIONS	5. TYPE OF REPORT & PERIOD COVERED: Technical rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Karen Kafadar	8. CONTRACT OR GRANT NUMBER(s): 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: Nov 79
	13. NUMBER OF PAGES: 34
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE AUG 18 1980 A

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

NA

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)

confidence
sampling
statistics
probability
simulation
Monte Carlo methods

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

A robust confidence interval using biweights for the case of five observations is proposed when the underlying distribution has somewhat heavier tails than the Gaussian. The distribution of a "t"-like statistic is approximated by a Student's t on the nominal four degrees of freedom using different scale factors which depend upon the value of the biweight weights. Results given by Monte Carlo simulations indicate that, even for very high coverage probabilities, the intervals proposed are highly efficient, in terms of the expected length of the confidence interval.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A ROBUST CONFIDENCE INTERVAL FOR SAMPLES OF FIVE OBSERVATIONS

by

Karen Kafadar

Princeton University

and

Oregon State University

Technical Report No. 154, 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. DAAG29-76-0298, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

ABSTRACT

A robust confidence interval using biweights for the case of five observations is proposed when the underlying distribution has somewhat heavier tails than the Gaussian. The distribution of a "t"-like statistic is approximated by a Student's t on the nominal four degrees of freedom using different scale factors which depend upon the value of the biweight weights. Results given by Monte Carlo simulations indicate that, even for very high coverage probabilities, the intervals proposed are highly efficient, in terms of the expected length of the confidence interval.

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 much helpful advice from Professors John W. Tukey and Peter Bloomfield during preparation of the dissertation and their numerous comments on early drafts of this paper.

0. INTRODUCTION

In an earlier report, the author [9] considered the use of the biweight in constructing a confidence interval for a sample of at least ten observations. Using the Student's t critical point on nine-tenths of the nominal degrees of freedom, it was found that the efficiency of a $100\cdot(1-\alpha)\%$ confidence interval in the Gaussian and in symmetric stretched-tailed situations exceeded 80% across a wide range for α. In this report, we bravely explore the performance of the biweight in a "t"-like statistic when we have only five observations. We are looking for good performance, not only in the (unlikely) event that our sample is truly Gaussian, but also if our sample comes from a population with somewhat heavier tails than the Gaussian.

Little is known about the results of robust procedures of the location problem alone on such small size samples. The Princeton Robustness Study [1] concluded that, in terms of 95% confidence intervals, the estimates could show considerable differences in non-Gaussian situations (Section 7B); their recommendation was a redescending Hampel-type estimator (Section 6L). Much of the literature on the interval problem for small samples has concentrated on the analytic distribution of Student's t statistic (e.g., [4], [7]). For more general stretched-tailed situations, several authors have shown that Student's t is highly conservative (e.g., [13], [15]). Except for specific underlying densities, a general solution to the interval problem has not been considered. The situation is particularly complicated by the facts that

(i) suspect "outliers," even in Gaussian samples, are not uncommon;

(ii) a 95% confidence interval for five Gaussian observations necessarily extends beyond the range of the data;

(iii) an extremely heavy-tailed situation offers just minimal amount of information required for a confidence interval, for, although the variance of a given M-estimate is finite, higher moments may not be.

While there exist many estimates to use in constructing robust confidence intervals, this report considers only the biweight in a "t"-like statistic, largely on the basis of its previous success in problems of interval estimation ([6], [9]), regression ([2]), and time series ([3]). This report is divided into three parts: Part A presents the results of biweight-"t" in the three sampling situations; Part B investigates a method to improve our estimate of the variance of the biweight via "compartmentalizing," and Part C offers conclusions and strategies for the case of five observations.

PART A. THE FAILURE OF "t"$_{bi}$ ON "UNUSUAL" SAMPLES.

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

For a definition of the biweight and its associated variance, the reader is referred to [12]; we mention here only the computational methods. 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

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

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 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) suggests 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, $T_{bi}^{(k)}$, $k \geq 1$, is found by

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

In determining an estimate of scale to use in (3), former studies (see, e.g., [1],[11]) 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 [11] showed that a more efficient scale estimate may be that using the functional form

$$s_{bi} = n^{1/2} \cdot (c_0 s^{(0)}) \cdot q_{411}(\{\tilde{u}_i\}) \qquad (4)$$

where

$$\tilde{u}_i = \frac{x_i - T^{(0)}}{c_0 \cdot s^{(0)}}$$

and

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

Here, as before, $T^{(0)}$ is the median of the sample, $s^{(0)}$ is the MAD, and c_0 is again chosen in order that $c_0 \cdot s^{(0)}$ is approximately the desired multiple of σ in the Gaussian case. (Since $s^{(0)} \simeq (2/3)\sigma$ for a Gaussian sample, we choose $c_0 = 9$ for this calculation.)

Finally, the denominator of our "t "$_{bi}$ statistic is given by S_{bi}, where S_{bi}^2 estimates the variance of T_{bi}. Huber [8] derives the theoretical asymptotic variance of T_{bi}, from which we may obtain a finite-sample approximation to it as

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

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

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

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)). Our biweight-"t" statistic then takes the form

$$"t"_{bi} = \frac{T_{bi}}{S_{bi}}.$$

2. <u>Results on samples of size five</u>.

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

- Gaussian
- One-Wild (4 observations from N(0,1);
  1 unidentified observation from N(0, 100))
- 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. The critical points of the distribution were all computed via a Monte Carlo swindle, the results of which may be found in [5]. There were 640 samples in the simulation for each sampling situation.

The success of biweight-"t" will be measured primarily in terms of "efficiency" of the expected confidence interval length (ECIL), i.e.,

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

where ECIL(α) was defined by Gross ([2]) as

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

and $\text{ECIL}_{min}(\alpha)$ is the "shortest" obtainable for the situation (see [9]). Furthermore, we shall be interested in approximating the distribution of biweight-"t" to a Student's t with some degrees of freedom, for practical purposes. Hence, we shall make the correspondence

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

When we examine the performance of biweight-"t" on samples of only five observations (Exhibit 1), we are initially disappointed with the results. Not only do we see low efficiencies in the lengths of the confidence intervals, but the matched degrees of freedom are unusually low. It appears that the numerator has extremely heavy tails; hence, $"t"_{bi}$ is matched to a Student's t with few degrees of freedom.

3. <u>"Unusual" Gaussian samples</u>.

In Exhibit 2 we consider the (swindled) estimate of one tail probability. Notice that these samples have been singled out because they resulted in unusually large estimates. All of these samples have the property that three of the five observations (just over half) are extremely close together, with the other two being far enough away that the bisquare function assigns them zero weight. Such Gaussian samples, although moderately rare, do occur with more than 2% frequency. In these cases, s_{bi} is sure to grossly underestimate σ, and $s^2_{bi} \ll \text{Var(numerator)}$, since the bisquare operates as for n=3 with extremely small variance. Any reasonable robust estimate of scale would perform likewise. Exhibit 2(b) presents location and scale estimates for more "typical" Gaussian samples. In these samples, s^2_{bi} is much closer in value to the usual sample s^2; hence, good performance in biweight-"t" is expected.

4. <u>Quantifying the behavior of "unusual" samples</u>.

If we can improve the estimate of

var(numerator)

in these problematic Gaussian samples, we may hope that a similar improvement may be used when the underlying distribution is not Gaussian. We therefore need a measure by which to classify the "unusual" samples. Returning to the formula (4) for s^2_{bi}, two possibilities for such a measure are suggested by:

a) $\sum_{i=1}^{n} w(u_i)$

or

b) $\sum_{i=1}^{n} \psi'(u_i)$

where

$$u_i = \frac{x_i - T_{bi}}{6\cdot(n^{1/2}S_{bi})}, \quad n=5.$$

We are particularly interested in distinguishing those samples for which one or more of the observations are far from the estimated center. This corresponds to $|u_i| > 1$, for which $\psi'(u_i)=w(u_i)=0$. Due to the monotonicity of $w(|u|)$, smaller values of the weight function always indicate increasingly greater distance. It appears likely that $\sum_{i=1}^{n} w(u_i)$ will be a more informative ancillary statistic than $\sum_{i=1}^{n} \psi'(u_i)$.

Exhibit 3 shows stem-and-leaf plots for the values of $\sum_{i=1}^{n} w(u_i)$ for the three sampling situations. The unusual Gaussian samples described above all fall among the samples for which $\sum_{i=1}^{n} w(u_i) \approx 2.94$. The majority of the samples have $\sum_{i=1}^{n} w(u_i) \approx 4.80$, for which S^2_{bi} performed adequately. The case where $\sum_{i=1}^{n} w(u_i) \approx 3.88$ corresponds to one observation being treated essentially as an outlier. The stem-and-leaf plots

suggest that the three cases can be specified in terms of a range of one unit in the value of

$$W = \sum_{i=1}^{n} w(u_i) \qquad (7)$$

Since we would like to choose the interval so as to most clearly differentiate among these samples, we choose end-points where the density of W is low:

$$\begin{aligned} &\text{two "false outliers":} && W \le 3.3 \\ &\text{one "false outlier":} && 3.3 < W \le 4.3 \qquad (8) \\ &\text{no "false outliers":} && W > 4.3. \end{aligned}$$

Here, "false" alludes to the fact that these observations, although some distance from the bulk of the sample, are nonetheless bonafide observations from the same distribution as the others. For the case of One-Wild where $3.3 < W \le 4.3$, the outlier does in fact usually correspond to the wild shot (from a N(0, 100) distribution). Henceforth, it will be convenient to analyze our results for n=5 not only by situation but by slice. A <u>slice</u> is defined by:

a) n, a given number of observations;

b) F, a distributional situation; (9)

c) a range of values, w_L and w_U, for which $w_L < W \le w_U$.

For a more detailed analysis of the effect of W on the biweight-"t" distribution, we generated nine slices of 600 samples each, where

a) n=5

b) F = Gaussian, One-Wild, or Slash

c) $W \le 3.3$, $3.3 < W \le 4.3$, $W > 4.3$.

Exhibit 4 tabulates the estimated frequencies for each slice, and the average values of the biweight and S^2_{bi} based on the 600 simulated samples. We see that a low-weight slice for n=5 is relatively infrequent, occurring in 2%-5% of all samples from our situations, yet the frequency is just large enough to produce the low efficiencies in the biweight-"t" intervals of Exhibit 1. Panel B of Exhibit 4 reveals that indeed the use of the biweight in the numerator of "t"$_{bi}$, despite its deflated scaling, is not the real problem, as its variance, even in the low-weight Gaussian samples, is only slightly more than twice the variance of the optimal mean. The biweight is a big success in the high-weight samples: notice that the variance of the optimal mean in the Gaussian situation is nearly attained, and that in all high-weight slices,

$$\text{ave}(\text{denominator of "}t\text{"}_{bi})^2 \simeq \text{var}(\text{numerator of "}t\text{"}_{bi}) \qquad (10)$$

In the medium-weight slice, Eqn. (10) already approximately holds for the more stretched-tailed distributions, but it is off by nearly a factor of 10 in the Gaussian situation. In order to achieve correspondingly good results for all medium-weight slices, it is likely that we will need to be conservative in some places.

Finally, the large discrepancy among situations in ave(S^2_{bi}) for the low-weight slices suggests that a deeper look at the behavior of these samples is required.

5. Digression: Granularity of the weight distribution.

It is worth commenting on the granularity of the distribution of $W = \sum_{i=1}^{n} w(u_i)$ for the three situations. This tendency is partly due to our scale estimate,

$$\hat{\sigma} = s_{bi} = n^{1/2} \cdot S_{bi}$$

in

$$w(u_i) = w((x_i - T_{bi})/(6\hat{\sigma}))$$

As a rather extreme case, consider the following estimate of $\hat{\sigma}$:

$$\hat{\sigma} = (1/6)\min(|x_1-x_3|, |x_2-x_3|, |x_4-x_3|, |x_5-x_3|)$$

where the sample $\underline{x}$ is assumed ordered $(x_1 \le x_2 \le \cdots \le x_5)$. Then

$$T = \frac{\sum x_i w(u_i)}{\sum w(u_i)} = x_3 \; ,$$

since $w(u_i)=0$ for all i except $i=3$, when $w(u_3)=1$. Hence, this functional form for $\hat{\sigma}$ will result in

$$W = \sum_{i=1}^{n} w(u_i) = 1 = \text{constant},$$

regardless of any further characteristics of the sample.

That this is a rather silly estimate for σ can be seen from the following two fabricated samples:

a) -1.6, -0.8, -0.6, 0.4, 1.0 ==> $\hat{\sigma}$ = 0.03;

b) -0.9, -0.8, -0.6, 0.4, 0.4 ==> $\hat{\sigma}$ = 0.03.

Nonetheless, the example does serve to indicate that the continuity of the density function of W is highly dependent upon choice of scale. It is quite possible that there exists a choice of scale for which W has a somewhat smoother density function. For reasonably efficient estimates of scale, however, its density is likely to have <u>modes</u> separated roughly by one unit (on the weight scale). The cutoff points we have selected in Equation (9) are likely to be satisfactory (i.e., to come at very low densities) for the weighting based on any reasonable scale estimate.

PART B: COMPARTMENTALIZING: SLICES.

6. A scaled biweight-"t" for slices.

Since our three weight classes in each situation vaguely represent the degree to which S^2_{bi} fails as an estimate of the variance of the biweight, a scaled version of "t"$_{bi}$, conditional on a given weight slice, might have a distribution which is more similar to a Student's t. That is, we would like to find a scale factor, K, such that

$$P\left(\frac{\text{"t"}_{bi}}{K} \geq a \mid w_L < W \leq w_U \,,\ n \right) \tag{11}$$

$$= P\left(\frac{T_{bi}}{KS_{bi}} \geq a \mid w_L < W \leq w_U \,,\ n \right) \doteq P(t_\nu > a)$$

where both K and ν may depend on W = Σ(weights) and on the sample size n.

One choice of K is suggested by the values in Exhibit 4(b). If we want to insist that $\nu = n-1$, and, in addition, that (10) hold approximately in all situations, we would choose our scale factors as follows:

	Gaussian	One-Wild	Slash	conservative K (max of three)
low W	15.49	8.45	2.21	15.49
medium W	4.36	1.05	0.90	4.36
high W	0.94	0.82	1.07	1.07

While these scale factors are all of the same order in the medium- and high-weight slices, clearly we may be much too conservative in the low-weight slice. Furthermore, it is not altogether certain that a matching of

$$\text{ave(denominator}^2) \simeq \text{var(numerator)}$$

will transform a biweight-"t" distribution into one from the Student's-t family.

An adaptive alternative may be based on dealing with the actual critical points from biweight-"t". Ideally,

$$\text{"t"}(\alpha_i)/t_\nu(\alpha_i) = \text{constant} \quad \text{for all } \alpha_i$$

or, equivalently,

$$y_i(\nu) = \log(\text{"t"}(\alpha_i)/t_\nu(\alpha_i)) = \text{constant} \quad \text{for all } \alpha_i$$

A least squares approach would minimize

$$\sum_{i=1}^{g} (y_i(\nu) - \text{constant})^2 \qquad (12)$$

whence

$$\text{constant} = \text{ave}(y_i(\nu)) = \bar{y}(\nu).$$

Then, minimizing (12) would be equivalent to

$$\min_{\nu} s^2(\nu)$$

where $s^2(\nu)$ is our sample variance formula

$$s^2(\nu) = \sum_{i=1}^{g} (y_i(\nu)-\bar{y}(\nu))^2/(g-1)$$

However, our $y_i(\nu)$ are not independent, and, even if they were, the usual sample s^2 would be vetoed on the basis of its overall nonrobustness. As we mentioned in [5], s^2_{bi} performs well with a sufficient number of observations [here we will use g=10, where g is the number of values of $y_i(\nu)$]. Moreover, Portnoy's results ([14]) indicate that using the redescending ψ-function may help us with at least one type of dependence among the observations. Let us therefore choose ν_0 by

$$\nu_0: \quad s^2_{bi}(\nu_0) \quad \text{is a minimum.} \qquad (13)$$

Secondly, we select a more conservative value for the constant by assigning

$$\log(\text{scale factor}) = \max_{1\le i\le g} y_i(\nu_0) \qquad (14)$$

7. Log(scale factors), by slice.

Exhibit 5 summarizes the degrees of freedom and log (scale factors) for the slices, both for ν_0=4 and for ν_0 chosen via (13). The closeness of our fit to a Student's t on ν degrees of freedom may be viewed graphically by plotting

$$\log((\text{scaled "t"}(\alpha_i))/t_\nu(\alpha_i)) \text{ vs } -\log(\alpha_i)$$

with one standard deviation "confidence limits" obtained from the curves

$\log((\text{scaled "t"}(\alpha_i) \pm \text{std error})/t_\nu(\alpha_i))$ vs $-\log(\alpha_i)$

Notice in these plots (Exhibit 6) that:

(1) a negative value of the ordinate indicates a conservative fit, and

(2) a negative slope suggests a larger value of ν.

The most successful approximation is the high-weight slice, for which

(biweight-"t")/0.95 *approximately* distributed as t_4 (c.f. Exhibit 6(a)).

For the medium-weight slice, a uniform rescaling of the biweight-"t" distribution more closely matches a Student's t on 3 d.f. One might argue that the Gaussian sample shows a "suspect" outlier, and a conservative, albeit wasteful, approach is to allow ourselves one fewer degree of freedom. There is only a small probability that we will waste this valid observation in the Gaussian case (0.028 from Exhibit 4). Of course, there is a much greater likelihood of obtaining a medium-weight One-Wild sample; in this case, inference based on four of the observations is a sensible procedure in the absence of knowledge of the kind of contamination. Exhibit 7(a) shows the relative improvement in comparing our scaled "t" points to Student's t on 3 d.f.

The scale factors for the low-weight slice, however, are still radically different. Not surprisingly, "t" needs to be adjusted more drastically when the underlying distribution is truly Gaussian. For a decent matching at the less extreme tail areas, a scaled "t" compared to 2 d.f. offers a possible approach (Exhibit 7(b)), but the approximation is still far from good.

Comparing the graphs for the two low-weight fits, one possible procedure is

for $.05 \le \alpha \le .0005$, compare "t"/21.0 to Student's t_2,

for $.0005 < \alpha \le .00001$, compare "t"/91.2 to Student's t_4.

It is, *however, worthwhile to characterize the differences in* the three low-weight classes. One difference is apparent from Exhibits 5 through 7: the scale factors and approximations for One-Wild and Slash are very similar to one another and each is considerably different from the Gaussian. If we had a method whereby we could discriminate the Gaussian samples from those whose underlying distribution has more stretched tails, we might be more successful in adaptively scaling biweight-"t". This idea is pursued in greater detail in [10].

PART C: CLOSE.

8. Conclusions for $n=5$.

The initial aim of this study, that of constructing a valid confidence interval for the center of a population for which we have only five observations, led to a more ambitious goal of characterizing the distribution of the biweight-"t" statistic. The case with $n=5$ is perhaps the most difficult of all: there are too many observations to develop an analytic solution, yet so few that the likelihood of obtaining potentially misleading samples cannot be ignored. In searching for a more complete description of the tail behavior of "t" on five observations, we discovered that a characterization based on the sum of the (biweight) weights offers a more satisfying approach.

When we can borrow information on width from several samples, we may compute both T_{bi} and S_{bi}^2 using a scale estimate pooled from all samples:

$$s_{pool} = (9s_{\rho})/\sqrt{Jn}\cdot\left[\frac{\sum_{i=1}^{Jn}\psi^2(u_i)}{\left(\sum_{i=1}^{Jn}\psi'(u_i)\right)\left(-1+\sum_{i=1}^{Jn}\psi'(u_i)\right)}\right]^{1/2}$$

$$s_{\rho} = \text{pooled MAD},\quad \underline{u} = (x_{11}-\dot{x}_1,\ldots,x_{nJ}-\dot{x}_J).$$

In this case, as Exhibit 8 reveals, biweight-"t" performs fairly well. (This Exhibit tabulates only the matched degrees of freedom and ECIL efficiencies for the sake of brevity; J refers to the number of samples from which scale information has been borrowed.) It is not always clear, however, if and when we may borrow. An unjustified usage of borrowing may be extremely misleading. In this regard, the sum of the weights may lend insight: an unexpectedly low value of W may caution us to treat this sample separately from the rest and not use it in borrowing width information. The borrowing issue plays a more important role in the two-sample problem [10].

In the absence of additional samples of, say, five observations, or any additional width information for our sample at hand, a conservative approach to the interval estimation problem for the single sample of five observations would be

A) For $\sum_{i=1}^{5} w(u_i) > 4.3$, use $t_4(\alpha)\cdot(0.95\cdot S_{bi})$
for the allowance in the confidence interval.
As we saw in Section 6, the resulting confidence interval performs very well;

B) For $3.3 < \Sigma w(u_i) \le 4.3$, use either $t_4(\alpha)\cdot(7.1\cdot S_{bi})$ or $t_3(\alpha)\cdot(4.8\cdot S_{bi})$ for the allowance;

C) For $\Sigma w(u_i) \le 3.3$,
   (i) for $.05 \le \alpha \le .001$, compare "t"/21.0 to Student's t_2;
   for $.001 < \alpha \le .00001$, compare "t"/91.2 to Student's t_4;
   (ii) consider the improvements, based on additional ancillary statistics, given in [10];
   (iii) pray for more information.

Dist'n.	Tail Pr.	Crit. Pt.	Stnd. Error	D.F.	ECIL	Efficiency
	0.00001	222.6	(8.206)	2.0	187.3	1.10
	0.000025	219.6	(7.654)	2.0	184.8	0.71
	0.00005	206.7	(7.596)	2.0	173.9	0.56
Dist'n.	0.00010	186.0	(7.397)	1.9	156.5	0.47
Gaussian	0.00050	118.1	(5.897)	1.7	99.34	0.53
	0.00100	86.79	(4.087)	1.6	73.03	0.68
	0.00500	26.93	(1.098)	1.4	22.66	2.92
	0.01000	13.20	(0.479)	1.6	11.11	8.04
	0.02500	4.325	(0.126)	2.0	3.639	41.16
	0.05000	2.650	(0.061)	2.3	2.230	64.64
	0.00001	133.8	(4.843)	2.6	169.9	1.33
	0.000025	120.5	(4.804)	2.2	153.0	1.04
	0.00005	109.0	(4.702)	2.0	138.4	0.89
Dist'n.	0.00010	95.77	(3.853)	2.0	121.6	0.81
One-Wild	0.00050	62.60	(2.633)	1.9	79.47	0.83
	0.00100	47.70	(2.231)	1.8	60.55	0.99
	0.00500	16.10	(0.626)	1.7	20.44	3.59
	0.01000	8.170	(0.281)	1.8	10.37	9.22
	0.02500	3.390	(0.092)	2.7	4.304	29.43
	0.05000	2.196	(0.047)	3.6	2.788	41.36
	0.00001	294.188	(13.037)	2.0	763.308	5.70
	0.000025	258.162	(12.597)	2.0	669.833	5.78
	0.00005	227.870	(12.035)	1.9	591.238	7.18
Dist'n.	0.00010	210.231	(10.899)	1.9	545.469	6.66
Slash	0.00050	102.154	(6.298)	1.7	265.051	5.44
	0.00100	53.148	(3.305)	1.8	137.901	5.88
	0.00500	15.011	(0.553)	1.7	38.947	7.42
	0.01000	8.820	(0.281)	1.8	22.886	9.17
	0.02500	3.954	(0.106)	2.2	10.258	11.58
	0.05000	2.472	(0.056)	2.7	6.415	10.06

Exhibit 1: Results on one-sample biweight - "t", n = 5

Exhibit 2

(A) $\hat{p}\{"t"_{bi} > 3.747\}$ for 14 "unusual" Gaussian samples.

Sample Number	x(1)	x(2)	x(3)	x(4)	x(5)	T_{bi}	S_{bi}	$\bar{x}$	S_{sample}	$\frac{S^2_{sample}}{S^2_{bi}}$	$\hat{p}$
25	-0.552	-0.434	-0.479	0.227	0.658	-0.504	0.024	-0.126	0.242	100.628	0.4437
59	-1.338	-1.292	-0.012	0.010	0.123	0.040	0.043	-0.502	0.333	59.957	0.4303
159	-0.093	-0.088	-0.070	0.132	1.667	-0.084	0.007	0.310	0.342	2439.455	0.4843
165	-0.958	-0.940	-0.930	-0.188	2.111	-0.942	0.008	-0.181	0.591	5041.095	0.4935
178	-1.800	-1.349	-0.056	-0.030	0.008	-0.026	0.019	-0.645	0.386	409.504	0.4755
299	-1.902	-0.540	0.540	0.551	0.607	0.566	0.021	-0.149	0.489	521.027	0.4799
328	-0.994	-0.816	-0.773	1.312	2.012	-0.860	0.070	0.143	0.629	81.376	0.4775
388	-0.480	-0.460	-0.448	0.608	1.400	-0.463	0.010	0.124	0.381	1527.198	0.4362
444	-0.270	-0.173	-0.158	0.834	1.014	-0.200	0.036	0.249	0.277	58.971	0.4271
511	-0.513	0.526	0.554	0.556	1.026	0.546	0.010	0.430	0.254	645.000	0.4678
515	-1.454	-1.442	-1.315	-0.115	0.765	-1.404	0.045	-0.712	0.446	98.448	0.4531
535	-0.898	-0.879	-0.775	0.396	1.187	-0.851	0.039	-0.194	0.422	117.617	0.4549
575	-0.265	0.517	0.544	0.552	2.094	0.533	0.011	0.688	0.384	1210.235	0.4652
604	-2.877	-2.346	-0.091	-0.013	0.039	-0.022	0.038	-1.058	0.640	278.543	0.4906
mean						-0.262	0.027	-0.116	0.415	899.220	0.4669
std err(mean)						0.162	0.005	0.130	0.035	360.291	0.0060

Exhibit 2 (continued)

(B) $\hat{p}\{"t"_{bi} > 3.747\}$ for 12 "typical" Gaussian samples.

sample number	x(1)	x(2)	x(3)	x(4)	x(5)	T_{bi}	S_{bi}	$\bar{x}$	S_{sample}	$\frac{S^2_{sample}}{S^2_{bi}}$	$\hat{p}$
606	-0.657	-0.022	0.217	0.345	2.733	0.131	0.404	0.523	0.579	2.052	0.0030
616	-0.472	0.361	0.615	0.693	0.741	0.568	0.137	0.388	0.225	2.679	0.1440
617	-0.321	0.234	0.247	0.703	0.892	0.356	0.224	0.351	0.211	0.886	0.0301
618	-0.457	0.633	0.852	1.070	1.114	0.700	0.293	0.640	0.290	0.979	0.0075
619	-0.633	-0.178	0.247	0.395	0.871	0.143	0.272	0.141	0.256	0.887	0.0114
621	-0.178	0.183	0.366	0.575	1.034	0.393	0.211	0.396	0.202	0.912	0.0333
622	-1.259	-0.574	0.129	0.612	0.912	-0.025	0.424	-0.036	0.395	0.870	0.0002
623	-1.273	-0.726	-0.686	-0.285	0.151	-0.565	0.251	-0.564	0.238	0.897	0.0177
624	-0.771	-0.507	-0.406	0.095	0.983	-0.143	0.322	-0.121	0.310	0.923	0.0035
627	-0.086	0.000	0.289	0.455	1.091	0.336	0.219	0.350	0.210	0.917	0.0333
629	-0.539	-0.538	-0.045	0.715	1.169	0.143	0.367	0.152	0.342	0.868	0.0010
631	-1.417	-0.996	-0.378	0.130	0.192	-0.487	0.340	-0.494	0.314	0.855	0.0022
mean						0.129	0.289	0.144	0.298	1.144	0.0240
std err (mean)						0.111	0.025	0.110	0.031	0.169	0.0120

Exhibit 3

Stem-and-leaf plots for $W=\Sigma w(u_i)$
for three sampling situations (n=5)

(i) 640 Gaussian samples:

```
29|4444444444444444444
30|
31|
32|
33|
34|
35|
36|
37|
38|8888888888888
39|0677
40|
41|15
42|347
43|07
44|136
45|56
46|57
47|01456777778888888889999999999999999999999999999999+16
48|00000000000000000000000000000000000000000000000000+99+
49|01244
```

(ii) 640 One-Wild samples:

```
29|444444444444444446
30|
31|
32|
33|
34|24
35|
36|
37|
38|36688888888888888888888888888888888888888888888888+99+
39|0122344555799
40|3446
41|3579
42|
43|0246
44|389
45|2235
46|1133357888
47|01122345566777777788888899999999999999999999999999+36
48|00000000000000000000000000000000000000000000000000+99+
```

Exhibit 3 (cont.)
Stem-and-leaf plots for $W=\Sigma w(u_i)$
for three sampling situations (n=5)

(iii) 640 Slash samples:

```
29|44444444444444444444444444444444444446
30|9
31|
32|
33|
34|
35|
36|
37|3
38|57788888888888888888888888888888888888888888888888+91
39|00223559 9
40|1233478
41|56
42|5
43|
44|3459
45|35579
46|3889
47|112334555667778888999999999999999999999999999999999+22
48|00000000000000000000000000000000000000000000000000+99+
49|4
```

(iv) Number of samples in three weight groups:

Dist'n	W≤3.3	3.3<W≤4.3	W>4.3
Gaussian	19	23	598
One-Wild	18	323	299
Slash	37	161	442

Exhibit 4
Information of slices for n=5

A) Relative frequencies (in %) of slices
(standard errors in parentheses)

	Gaussian	One-Wild	Slash
$W \leq 3.3$	2.43 (0.10)	2.80 (0.11)	4.57 (0.18)
$3.3 < W \leq 4.3$	2.76 (0.10)	45.57 (0.34)	21.14 (0.36)
$W > 4.3$	94.81 (0.14)	51.63 (0.34)	74.29 (0.38)

B) Some summary values on T_{bi} and S^2_{bi}, by slice

		Gaussian	One-Wild	Slash
$W \leq 3.3$	var(T_{bi})	0.480 (.014)	0.500 (.011)	1.272 (.065)
	ave(S^2_{bi})	0.002 (.0001)	0.007 (.0001)	0.260 (.032)
$3.3 < W \leq 4.3$	var(T_{bi})	0.380 (.005)	0.253 (.001)	2.242 (.220)
	ave(S^2_{bi})	0.020 (.015)	0.230 (.002)	2.771 (.306)
$W > 4.3$	var(T_{bi})	0.203 (.004)	0.928 (.033)	9.450 (5.050)
	ave(S^2_{bi})	0.228 (.008)	1.219 (.062)	8.300 (2.647)

Exhibit 5
Log(scale factors), by slice

A) Fitted log(scale factors) (=log(K)) and degrees of freedom (=ν):

		Gaussian	One-Wild	Slash
Low	ν	2	2	3
Weight	log(K)	1.59	1.22	1.53
Medium	ν	3	4	3
Weight	log(K)	0.68	0.28	0.17
High	ν	5	5	4
Weight	log(K)	0.064	-0.035	-0.10

B) Log(scale factors) for degrees of freedom = 4:

		Gaussian	One-Wild	Slash
Low	ν=4:	1.79	1.96	1.71
Medium	ν=4:	0.85	0.28	0.34
High	ν=4:	-0.016	-0.12	-0.10

C) Log(scale factors) for ν=-1+[-1+Σweights]:

		Gaussian	One-Wild	Slash
Low	ν=2:	1.32	1.22	1.18
Medium	ν=3:	0.68	0.13	0.13
High	ν=4:	-0.016	-0.12	-0.10

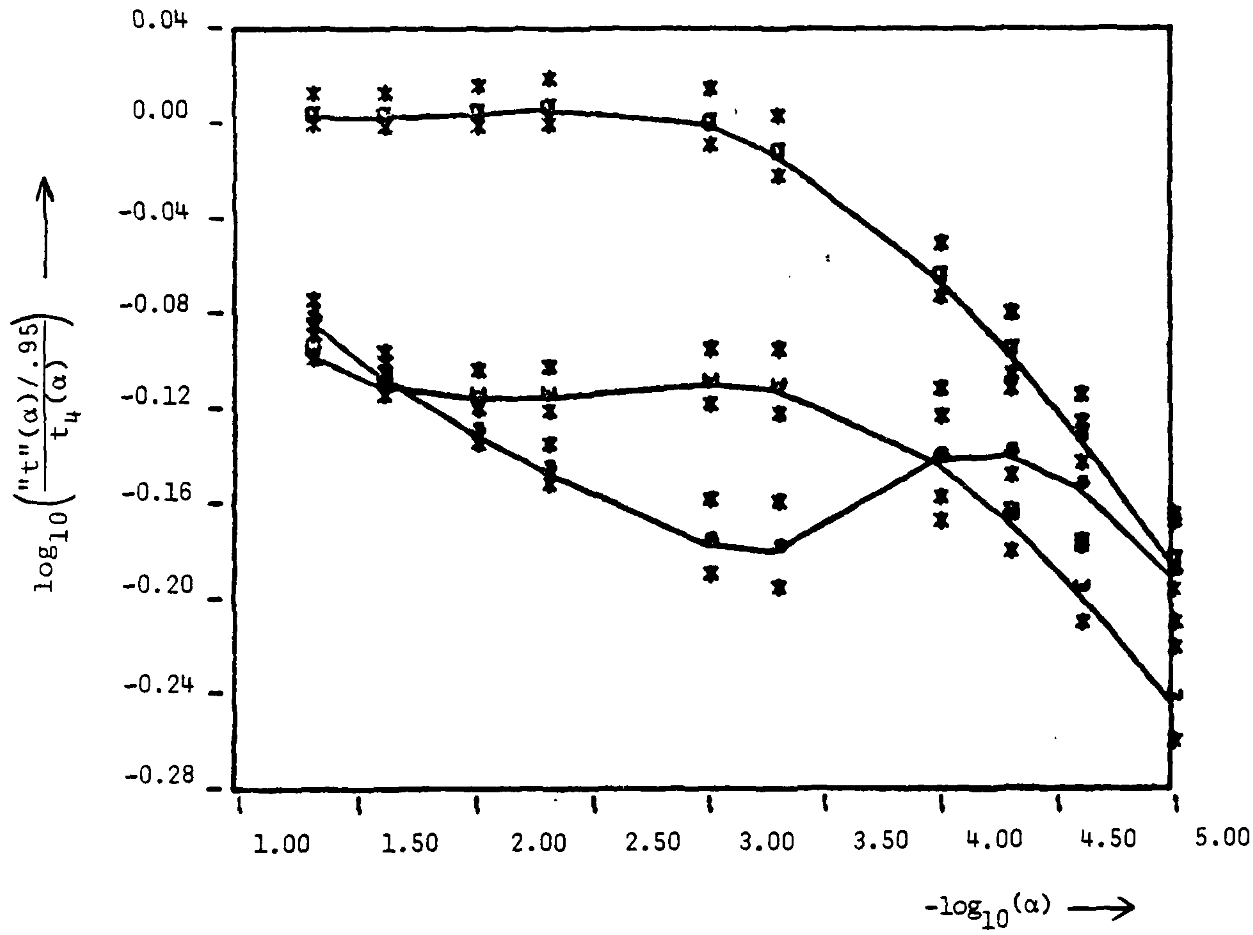


Exhibit 6(a). Plot of $\log_{10}\left(\frac{"t"(\alpha)/.95}{t_4(\alpha)}\right)$ vs. $-\log_{10}(\alpha)$

on high-weight slices, n=5.

(g=Gaussian; w=One-Wild; s=Slash;
including 1 std. dev. in "t"(α))

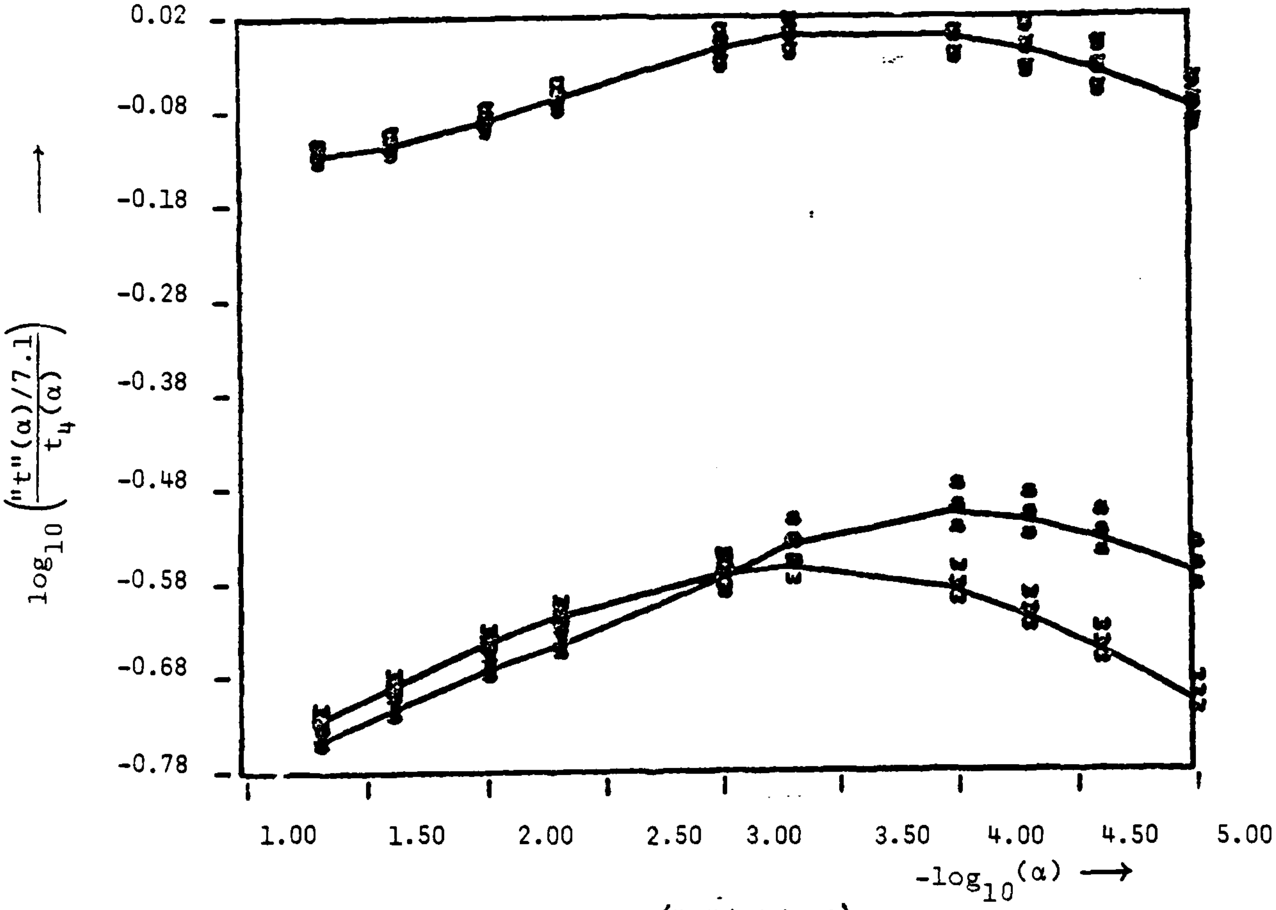
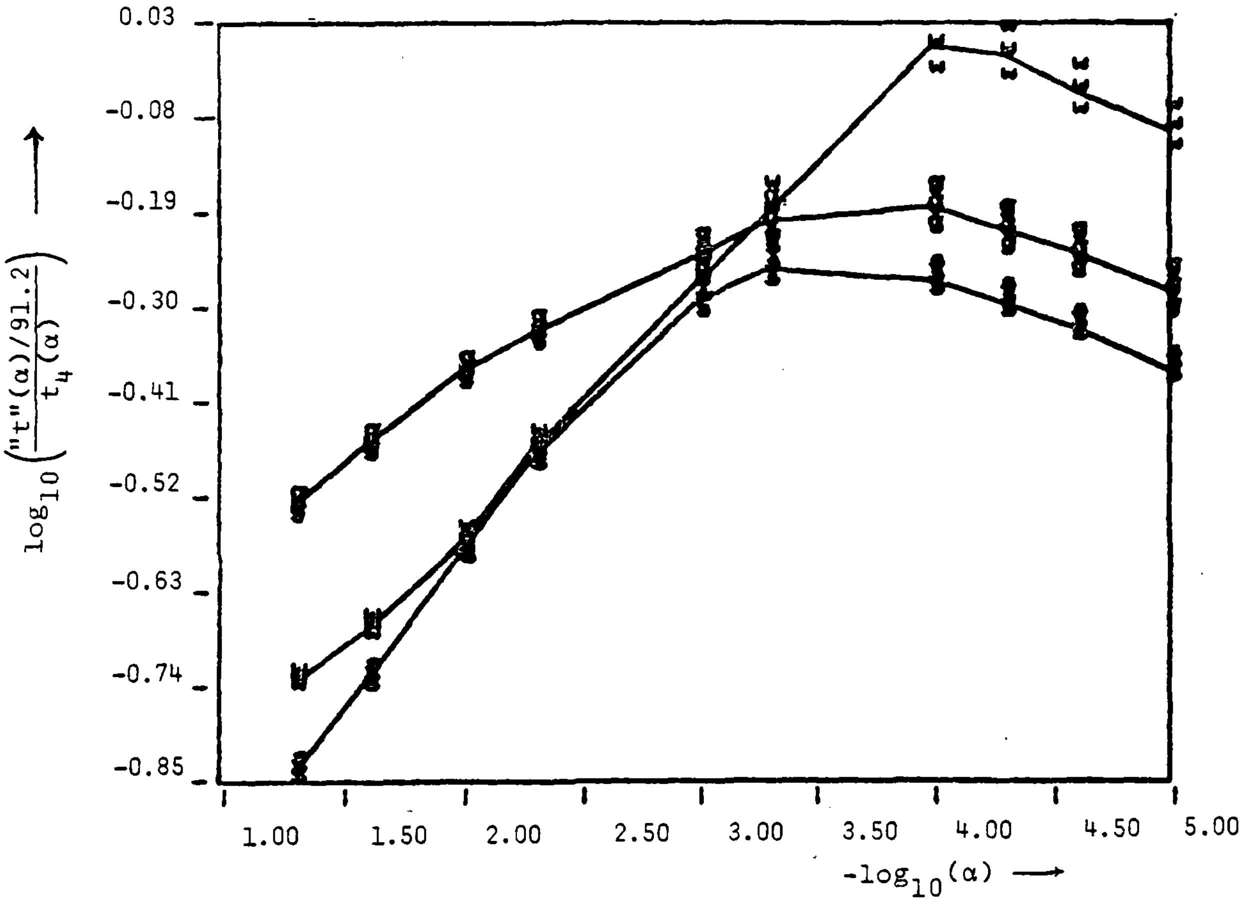


Exhibit 6(b). Plot of $\log_{10}\left(\frac{"t"(\alpha)/7.1}{t_4(\alpha)}\right)$ vs. $-\log_{10}(\alpha)$

on medium-weight slices, n=5.

(g=Gaussian; w=One-Wild; s=Slash;

including 1 std. dev. in "t"(α))



<u>Exhibit 6(c)</u>. Plot of $\log_{10}\left(\frac{"t"(\alpha)/91.2}{t_4(\alpha)}\right)$ vs. $-\log_{10}(\alpha)$

on low-weight slices, n=5.

(g=Gaussian; w=One-Wild; s=Slash;

including 1 std. dev. in "t"(α))

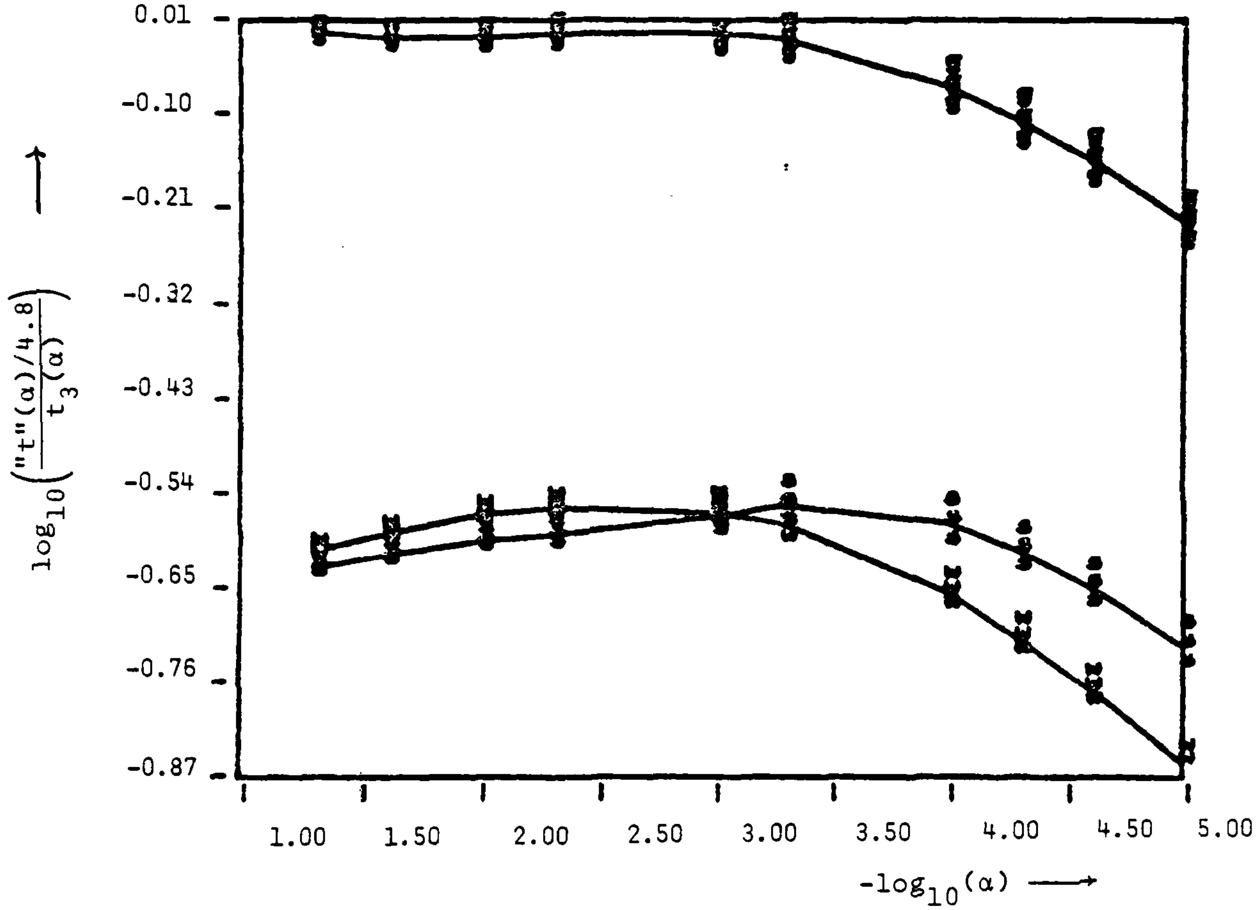
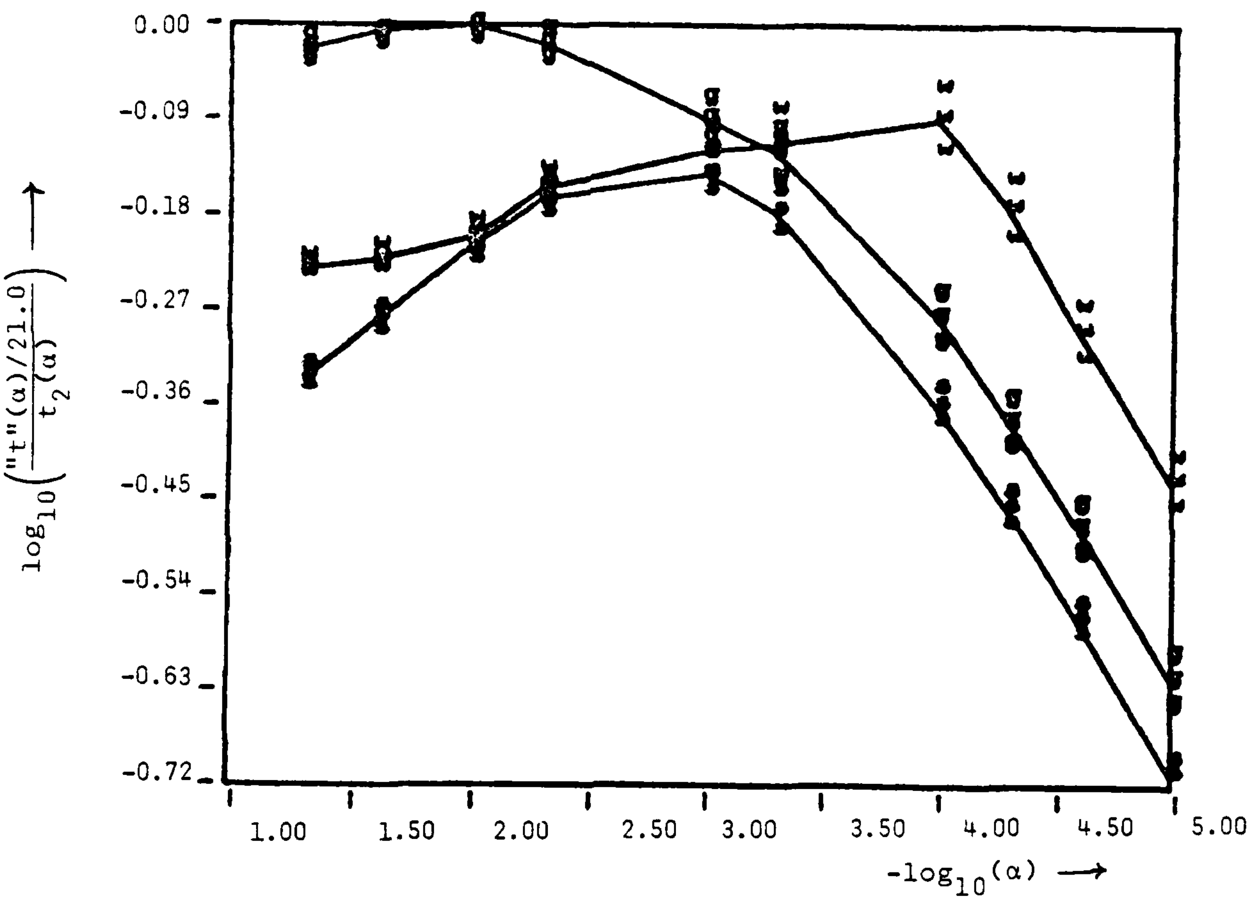


Exhibit 7(a). Plot of $\log_{10}\left(\frac{"t"(\alpha)/4.8}{t_3(\alpha)}\right)$ vs. $-\log_{10}(\alpha)$ on medium-weight slices, n=5. (g=Gaussian; w=One-Wild; s=Slash; including 1 std. dev. in "t"(α))



<u>Exhibit 7(b)</u>. Plot of $\log_{10}\left(\frac{"t"(\alpha)/21.0}{t_2(\alpha)}\right)$ vs. $-\log_{10}(\alpha)$ on low-weight slices, n=5.
(g=Gaussian; w=One-Wild; s=Slash; including 1 std. dev. in "t"(α))

Exhibit 8

Biweight-"t" on n=5: Borrowed numerators and denominators, using pooled ASYMV.

#borrowed	2		3		4		5	
α	d.f.	eff.	d.f.	eff.	d.f.	eff.	d.f.	eff.
				Gaussian				
.00001	6.2	97.0	8.5	82.0	9.0	85.1	19.1	86.8
.00005	6.2	96.2	8.4	84.8	9.5	88.6	18.6	85.7
.0001	6.0	96.1	8.5	86.5	9.3	91.1	18.5	85.8
.0005	5.9	98.4	9.5	90.8	11.7	92.0	19.3	87.7
.001	5.5	99.3	10.2	99.0	13.1	98.5	20.3	89.1
.005	5.6	101.6	11.6	101.1	16.2	107.6	26.3	93.2
.01	5.9	115.7	12.5	103.7	18.4	110.4	33.7	95.0
.025	6.5	127.4	14.5	103.7	25.1	113.7	82.0	97.5
.05	7.0	133.4	18.4	105.7	47.4	116.2	∞	99.4
				One-Wild				
.00001	5.2	58.1	8.9	70.9	9.3	92.0	24.4	85.1
.00005	5.8	60.5	9.0	70.1	9.3	91.5	28.0	84.4
.0001	6.0	63.1	9.0	66.5	9.7	91.1	30.4	84.2
.0005	6.2	64.0	8.8	64.0	13.4	90.2	42.3	83.3
.001	5.8	68.1	8.3	62.4	16.6	94.7	55.3	83.3
.005	5.8	75.2	12.9	69.2	34.1	96.1	∞	83.2
.01	6.6	81.2	16.7	71.0	76.0	96.9	∞	82.8
.025	8.5	83.5	32.7	71.5	∞	96.6	∞	82.1
.05	11.6	83.1	∞	70.9	∞	95.2	∞	80.8
				Slash				
.00001	7.8	41.2	10.0	48.3	12.2	50.1	11.2	49.3
.00005	7.7	40.4	10.0	45.4	12.1	52.2	10.9	48.8
.0001	7.5	40.0	9.8	46..1	11.3	51.1	11.0	48.2
.0005	7.5	42.1	10.4	49.9	11.5	49.2	11.7	47.6
.001	8.0	44.2	10.6	52.2	12.0	48.0	11.9	48.5
.005	8.2	50.0	11.1	54.3	12.4	49.9	12.1	49.1
.01	8.4	51.5	11.2	58.1	16.4	51.8	17.5	50.8
.025	11.5	52.3	19.0	60.7	54.1	52.2	∞	51.9
.05	23.1	52.3	202.4	68.6	∞	55.5	∞	53.0

REFERENCES

[1] Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, New Jersey.

[2] Denby, L. and Larsen, W.A. (1977). Robust Regression Estimators Compared via Monte Carlo. *Comm. in Statist. A6*, 335-362.

[3] Denby, L. and Martin, R.D. (1979). Robust Estimation of the First-Order Autoregressive Parameter. *J. Amer. Statist. Assoc., 74*, No. 365, 140-146.

[4] Geary, R.C. (1936). The Distribution of "Student's" Ratio for Non-normal Samples. Supplement to *J. Royal Statist. Soc., 2*, 178-184.

[5] Gross, A.M. (1973). A Monte Carlo Swindle for Estimators of Location. *Appl. Statist., 22*, No. 3, 347-353.

[6] __________ (1976). Confidence Interval Robustness with Long-Tailed Symmetric Distributions. *J. Amer. Statist. Assoc., 71*, 409-417.

[7] Hotelling, H. (1961). The Behavior of Standard Statistical Tests under Non-Standard Conditions. *Proceedings of the Fourth Berkeley Symposium I*, 319-361.

[8] Huber, P. (1977). *Robust Statistical Procedures*. Philadelphia: Society for Regional and Applied Mathematics (Regional Conference Series in Applied Mathematics, 27).

[9] Kafadar, K. (1979). A Biweight Approach to the One-Sample Problem. Tech. Report No. 151, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

[10] __________ (1979). Robust Confidence Intervals for the One- and Two-Sample Problems. Ph.D. Dissertation, Department of Statistics, Princeton University, Princeton, New Jersey.

[11] Lax, D. (1975). An Interim Report of a Monte Carlo Study of Robust Estimates of Width. Tech. Report No. 93, Series 2, Department of Statistics, Princeton University, Princeton, New Jersey.

[12] Mosteller, F. and Tukey, J.W. (1972). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, Mass.

[13] Pearson, E.S., assisted by Adyanthaya, N.K. (1929). The Distribution of Frequency Constants in Small Samples from Nonnormal Symmetrical and Skew Populations. *Biometrika, 21*, 254-289.

[14] Portnoy, S. (1977). Robust Estimation in Dependent Situations. *Ann. Statist., 5*, 22-43.

[15] Tukey, J.W. and McLaughlin, D. (1963). Less Vulnerable Confidence and Significance Procedures for Location Based on a Single Sample: Trimming/Winsorization. 35 *Sankhyo, Series A*, Part 3, 331-352.