Section I. Introduction

We have implemented a SAS-76 procedure, WILC0X, which obtains a distribution-free estimate with confidence intervals for the ratio of two random variables. This paper discusses the options and parameters of the procedure WILC0X. Results from a simulation study comparing the small sample properties of the distribution-free estimator and other estimators are presented, along with an application to environmental impact assessment.

Suppose \((X_i, Y_i)\) \(i = 1, 2 \ldots n\) are \(n\) matched pairs. The ratio \(p\) is defined as

\[
E[Y] = pE[X], \quad p > 0 .
\]

(1)

This type of model is common in bioassay (Finney 1952). Bennett (1965) obtained a distribution-free estimate of \(p\) as follows:

For a given value of \(p\) define the \(n\) pairs \((Y_i, U_i(p))\), \(i = 1, 2 \ldots n\) where

\[
U_i(p) = pX_i .
\]

(2)

Furthermore, let \(T(p)\) be the Wilcoxon signed rank statistic based on \((Y_i, U_i(n))\). If \(p^*\) is the true value of \(p\), then the expected value of \(T(p^*)\) is given by

\[
E[T(p^*)] = \frac{4n(n+1)}{3} .
\]

(3)

It is readily shown (Sen 1963) that \(T(p)\) is a monotone decreasing function of \(p\). Bennett defines the estimate \(p^*\) as that value of \(p\) for which \(T(p)\) is closest to its expected value, i.e.,

\[
\left| T(p^*) - \frac{4n(n+1)}{3} \right| = \inf_{\rho} \left| T(\rho) - \frac{4n(n+1)}{3} \right| .
\]

(4)

The value of \(p^*\) is obtained as the root of

\[
F = T(\rho) - E[T(p^*)] .
\]

(5)

As noted earlier, \(T(\rho)\) is a monotone decreasing function of \(\rho\) and hence any root finding technique will suffice. In WILC0X the modified regula falsi method is used (Conte and Boor 1965). It can be shown (using arguments similar to Sen (1963) that \(\rho^*\) is a consistent estimator of \(p\) and is asymptotically normally distributed. Bennett defined \((\hat{\rho}, \rho^*)\), the limits of the 100\((1-\epsilon)\) percent confidence interval for \(\rho\) as follows:

\[
\frac{1}{2} < \hat{\rho}_c < \frac{1}{2}, \quad \rho \leq \hat{\rho}_e \leq \rho .
\]

(6)

\(\hat{\rho}_c\) and \(\hat{\rho}_e\) have been tabulated for different sample sizes and different \(\epsilon\)'s (Owen 1962). The limits are then defined as:

\[
\begin{align*}
T(\rho) - \hat{\rho}_c &= \inf_{\rho} \left| T(\rho) - \hat{\rho}_c \right| , \\
T(\rho) - \hat{\rho}_e &= \inf_{\rho} \left| T(\rho) - \hat{\rho}_e \right| .
\end{align*}
\]

(7a, 7b)

\(\rho^*\) and \(\rho\) can be obtained using the modified regula falsi method as before. It is well known that the statistic

\[
T_+ = \frac{T - n(n+1)/4}{\frac{n(n+1)(2n+1)/24}^{1/2}}
\]

(8)

has an asymptotic \(N(0,1)\) distribution. Hence, one can approximate \(\hat{\rho}_c\) and \(\hat{\rho}_e\) in equation (7) by

\[
\begin{align*}
\hat{\rho}_c &= n(n+1)/4 Z_{\rho/2} \left( \frac{n(n+1)(2n+1)/24}^{1/2} \right) , \\
\hat{\rho}_e &= n(n+1)/4 Z_{\rho/2} \left( \frac{n(n+1)(2n+1)/24}^{1/2} \right) ,
\end{align*}
\]

(9)

where \(Z_{\rho/2}\) is obtained from the \(N(0,1)\).

In the following sections the distribution-free estimator will be referred to as Bennett estimator.

Section II. Description of WILC0X

Besides the Bennett estimator discussed earlier, procedure WILC0X computes the following three estimators of \(p\):

\[
\begin{align*}
T_1 &= \text{ratio of means} = \frac{\bar{Y}}{\bar{X}} , \\
T_2 &= \text{mean of the ratios} = \frac{1}{n} \sum Y_i X_i = \frac{1}{n} \sum r_i , \\
T_3 &= \text{median of the ratios} (r_1, r_2, \ldots, r_n) .
\end{align*}
\]

In this section the various options and parameters of WILC0X are described.

(a) Procedure Statement:

\[\text{PROC WILC0X options_and_parameters;}\]

(b) Variables List:

A VAR card must accompany the PROC statement. The variables must be specified in pairs, e.g., VAR X Y A B;
(note that the model \( Y = pX \)). If an odd number of variables are specified, the last variable is ignored and a warning is printed in the SAS log.

(c) By list:

The form of this is the same as that in the SAS-76 User's Guide, p. 55 (Barr et al. 1976).

(d) Missing values:

If either or both the variables being processed are missing, the observation will be ignored. This will not affect the remaining pairs of variables specified in the VAR card.

(e) Options and Parameters:

RATIO If this parameter is specified the ratio estimators \( T_2, T_3, T_4 \) and the Bennett estimator will be obtained. Confidence intervals for the Bennett estimator will be computed.

ZERO WILCOX will ignore an observation, if either or both the variables being processed have a value of 0.0. By specifying this option, these observations will be included in the analysis.

ALPHA = The value specified must be some number between 0 and 1. The default value is 0.1. This is used to obtain \( T_E \) and \( T_{EC} \) using equation (9). If LOWER and UPPER are specified, the ALPHA parameter is ignored. The percentile of the \( N(0,1) \) is computed using a rational approximation (Abramowitz and Stegun 1970).

LOWER = The user can specify \( T_{EC} \) by specifying this parameter.

UPPER = The user can specify \( T_{EC} \) by specifying this parameter.

N = The procedure will bypass all computations, if the sample size is less than or equal to N. The default value for N is 5.

(f) In Fig. 1, a sample output is presented. The output is self-explanatory.

Section III. Some Simulation Results

In this section, the results of a simulation study comparing the small sample properties of the Bennett estimator and the \( T_E \) estimator (defined below) are presented. The results discussed here are only for a sample of size 20 and are based on 2000 replicates.

(a) Model:

In Fig. 2, the model for which the simulation experiment was conducted is shown. A random variable \( X_i \) is generated from a lognormal distribution with parameters \( \mu_i \) and \( \sigma_i^2 \). Two additional random variables \( X_j \) and \( Y_i \) are drawn from lognormal distributions with parameters \( (\mu_X = \mu_Y, \sigma^2_X) \) and \( (\mu_Y = \mu_Y, \sigma^2_Y) \), respectively.

The parameter to be estimated is the true ratio \( p \). The underlying distribution was chosen as the lognormal since it is common practice in environmental models to assume a lognormal distribution. Unfortunately, more often than not, the data do not appear to have come from a lognormal model. Hence, the interest in the Bennett estimator.

The Bennett estimator is compared against estimator \( T_E \) defined as

\[
\ln T_5 = \ln Y_i - \ln X_i. \tag{10}
\]

Note that in \( T_5 \) is the maximum likelihood estimator of \( \ln p \) for the model described above. In Fig. 3, the design for the simulation is presented. The full experiment given in (iii) of Fig. 3 was used in the study. Since specifying the parameter \( \sigma^2 \) is equivalent to specifying the coefficient of variation \( y \), the coefficient of variation was chosen as the design parameter.

(b) Efficiency and Bias:

The relative efficiency and average bias based on 2000 replicates and a sample size of 20 is reported in Tables 1 and 2. Relative efficiency is defined as

\[
\text{relative efficiency} = \frac{\text{variance} (T_E)}{\text{variance} (\text{Bennett estimator})}. \tag{11}
\]

Table 1 shows that the relative efficiency varies between 0.67 and 0.77. For the special case of \( p = 1 \) the efficiency is generally greater than 0.7.

In Table 2 the average biases for the Bennett estimator and the \( T_E \) estimator are summarized. It is well known that the \( T_E \) estimator is a biased estimator of \( p \) and the bias increases with increasing \( p \) and increasing \( y \), where \( y \) is the coefficient of variation. Table 2 shows that the Bennett estimator follows the same pattern. However, the bias is generally larger for the Bennett estimator.
(c) Robustness:

Another property of interest is the robustness, or lack of it, with respect to outliers. The Bennett estimator appears to be robust with respect to outliers. In Figs. 4 and 5 the stylized sensitivity curves (S.S. curves) (Andrews et al., 1972) for the Bennett estimator and the $T_s$ estimator are shown for the case of $p = 1$, and different values of $\gamma$. These curves were obtained as follows:

The $X_i$'s in the model described earlier were replaced by the expected order statistics from a sample of size 20. The $Y_i$'s $i = 1, 2, \ldots, 19$ were replaced by the expected order statistics from a sample of size 19, with the appropriate translation as specified by $p$, the true ratio and $\gamma$, the coefficient of variation. The 20th point $(Y_{20})$ was varied such that $r = Y_{20}/X_{20}$ increased from 0.1 to 10. For various values of $r$ the estimates $p(r)$ of $p$, the Bennett estimator and $T_s(r)$ of $T_s$ were obtained. These estimates were then multiplied by the sample size (20) and plotted against $r$. From Fig. 4 one observes that the Bennett estimator shows a sharp increase as $r$ increases beyond 1. However, as $r$ continues to increase, the estimator remains constant, i.e., S.S. curve is bounded in the tail. This indicates that an outlier has only a limited effect on the estimator (Andrews et al., 1972).

In Fig. 5 the S.S. curves for the $T_s$ estimator are shown. One observes that as $r$ increases, the estimator continues to increase in value, i.e., the tail of the S.S. curve is not bounded. This indicates that as the magnitude of the outlier $Y_{20}$ increases, the $T_s$ estimator moves farther and farther away from the "true" value.

(d) General Remarks:

The Bennett estimator appears to be a useful estimator of the ratio of two random variables since:

(a) it is easy to compute,
(b) since the Bennett estimator is distribution free, it can be used when unable to verify the distributional assumptions,
(c) while it has a bias, the bias is small, and
(d) it is robust with respect to outliers.

Section IV. An Application

In environmental impact assessment, one is interested in detecting any significant change in a biological parameter after intervention. The intervention could be in the form of hot water discharges from a nuclear power plant, a refinery, etc. A monitoring program is instituted which measures important biological parameters in the stressed and in unstressed (control) areas, both before and during the intervention. If there has been a significant impact, one would expect to observe a change in the relative behavior of the parameter in the control and stressed areas.

The ratio method can be used to detect such changes. One obtains estimates of the ratio of the stressed area to the control area over a period of several years. If a detectable impact has occurred, the estimated ratios after the intervention would be different from those during the pre-stress period.

Consider as an example the monitoring of zooplankton in a waterbody into which a nuclear station discharges heated water. The density of zooplankton was monitored for several years during the months of June through October at several stations. One of the stations (601) was far upstream and was used as control. In Fig. 6 the Bennett estimator for the various stations is plotted over the years. Here the model is zooplankton at station (i) = $p$ (zooplankton at station 601). Stations numbered 605 and 608 are in the thermal plume while 602 is at the water intake structure. The other two locations are far removed from the power plant. Observe that in the pre-operational years (1967-1972) the values are clustered together while in the operational years (1973-1975) they show a greater spread. It is interesting to note that in 1975 the three stressed stations have a smaller value for the ratio than the unstressed stations. Subsequent analysis of the data for pattern and structure showed that the stressed stations formed a group.

Section V: Conclusions

In this paper a new SAS-76 procedure WILCOX for obtaining distribution-free estimates (Bennett estimator) of ratio of two random variables was described. The small sample properties of this estimator was discussed in relation to the maximum likelihood estimator for a model based on the lognormal distribution. The small sample efficiency of this estimator was about 70% for a sample of size 20. The robustness of the Bennett estimator was demonstrated via the stylized sensitivity curves. An application to an environmental impact assessment was presented.

Acknowledgment

The authors wish to express their appreciation to our colleagues J. J. Beauchamp and R. J. Olson for their constructive criticism on the article and to C. L. Bayne and E. L. Miller for their review suggestions.
References


Footnotes

1 Research sponsored in part by the Energy Research and Development Administration under contract with Union Carbide Corporation and in part by the U.S. Nuclear Regulatory Commission Office of Nuclear Reactor Regulation under Interagency Agreement ERDA 40-554-75.

2 Publication No. 1013, Environmental Science Division, Oak Ridge National Laboratory.

3 Environmental Data Resources Group, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830.

Fig. 1. Sample output from PROC WILCOX.

Fig. 2. Modal for simulation experiment to investigate small sample properties of Bennett estimator.

Fig. 3. Design for the Simulation Experiment for the Model in Fig. 2.
Table 1. Relative efficiency$^1$ of Bennett estimator for sample size of 20$^2$

<table>
<thead>
<tr>
<th>Ratio (ρ)</th>
<th>0.2</th>
<th>0.6</th>
<th>1.0</th>
<th>1.4</th>
<th>1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.75</td>
<td>0.71</td>
<td>0.72</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>1.3</td>
<td>0.71</td>
<td>0.71</td>
<td>0.73</td>
<td>0.73</td>
<td>0.69</td>
</tr>
<tr>
<td>1.5</td>
<td>0.72</td>
<td>0.74</td>
<td>0.70</td>
<td>0.69</td>
<td>0.67</td>
</tr>
<tr>
<td>1.7</td>
<td>0.75</td>
<td>0.71</td>
<td>0.71</td>
<td>0.67</td>
<td>0.70</td>
</tr>
<tr>
<td>1.9</td>
<td>0.71</td>
<td>0.71</td>
<td>0.72</td>
<td>0.71</td>
<td>0.70</td>
</tr>
</tbody>
</table>

$^1$ Relative efficiency = $\text{var}(T_5)/\text{var(}\text{Bennett estimator})$ where $T_5 = T_5 - \bar{T}_5$.

$^2$ These results are based on 2000 replications.

Table 2. Estimated average bias$^{(1)}$ of Bennett estimator and $T_5^{(2)}$ estimator for a sample size of 20

<table>
<thead>
<tr>
<th>Ratio (ρ)</th>
<th>0.2</th>
<th>0.6</th>
<th>1.0</th>
<th>1.4</th>
<th>1.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>0.0065$^{(3)}$</td>
<td>0.00193</td>
<td>0.0050</td>
<td>0.00389</td>
<td>0.0037</td>
</tr>
<tr>
<td>1.3</td>
<td>-0.0007</td>
<td>0.00023</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>1.5</td>
<td>0.0025</td>
<td>0.00154</td>
<td>0.00036</td>
<td>0.00035</td>
<td>0.00035</td>
</tr>
<tr>
<td>1.7</td>
<td>0.0042</td>
<td>0.00317</td>
<td>0.00270</td>
<td>0.00264</td>
<td>0.00262</td>
</tr>
<tr>
<td>1.9</td>
<td>0.0056</td>
<td>0.00356</td>
<td>0.00264</td>
<td>0.00255</td>
<td>0.00242</td>
</tr>
</tbody>
</table>

$^{(1)}$ Based on 2000 replications.

$^{(3)}$ $T_5 = T_5 - \bar{T}_5$.

$^{(1)}$ Average bias for Bennett estimator.

$^{(2)}$ Average bias for $T_5$.

Fig. 4. Stylized sensitivity curves for Bennett estimator with $\sigma = 1$ and $\gamma = 0.2(0.4)1.8$.

Fig. 5. Stylized sensitivity curves for $T_5$ estimator with $\sigma = 1$ and $\gamma = 0.2(0.4)1.8$.
Fig. 6. Bennett estimator for zooplankton.