COMPARISON OF PROCEDURES FOR TESTING
THE EQUALITY OF SURVIVAL DISTRIBUTIONS

by

HARRY SUE WAH | JOE

B.Sc., The University of Victoria, 1978

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
(Department of Mathematics)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1979

© Harry Sue Wah Joe, 1979

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study.

I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date Sept. 11, 1979

ABSTRACT

A common problem in life testing is assessing whether two survival distributions are identical. Sequential procedures based on the Mantel-Haenszel statistic (Muenz, Green and Byar, 1977), the Wilcoxon statistic (Davis, 1978) and the Savage statistic (Koziol and Petkau, 1978) have been proposed for testing the equality of two survival distributions. These procedures, and their fixed point censoring counterparts, are numerically compared on the basis of power under Lehmann alternatives and expected proportion sampled. Among the sequential procedures, those based on the Wilcoxon and Savage statistics seem preferable to that based on the Mantel-Haenszel statistic. Early decision rules are introduced and shown to be of value if used in conjunction with either the sequential or fixed point censoring procedures. Finite sample corrections to the asymptotic critical values of the sequential procedure proposed by Koziol and Petkau are presented.

TABLE OF CONTENTS

ABSTRACT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	v
LIST OF FIGURES	vi
ACKNOWLEDGEMENT	vii
 Chapter 1 INTRODUCTION	 1
Chapter 2 DESCRIPTION OF PROCEDURES	6
2.1 INTRODUCTION	6
2.2 PROCEDURES BASED ON THE SAVAGE STATISTIC	6
2.3 PROCEDURES BASED ON THE WILCOXON STATISTIC	8
2.4 PROCEDURES BASED ON THE MANTEL-HAENSZEL STATISTIC ...	9
2.5 THE MODIFIED MGB SEQUENTIAL PROCEDURE	11
Chapter 3 NULL DISTRIBUTION RESULTS	13
3.1 INTRODUCTION	13
3.2 THE KP PROCEDURE, WITH EQUAL SAMPLE SIZES	13
3.3 THE DAVIS, MGB AND MODIFIED MGB PROCEDURES, WITH EQUAL SAMPLE SIZES	27
3.4 THE SEQUENTIAL PROCEDURES, WITH UNEQUAL SAMPLE SIZES..	29
3.5 THE FIXED POINT CENSORING PROCEDURES, WITH EQUAL SAMPLE SIZES	33
Chapter 4 COMPARISON OF PROCEDURES	35

TABLE OF CONTENTS

Chapter 5	EARLY DECISION RULES	44
Chapter 6	EXAMPLES	50
6.1	THE FIRST EXAMPLE	50
6.2	THE SECOND EXAMPLE	52
Chapter 7	DISCUSSION	55
BIBLIOGRAPHY		57

LIST OF TABLES

TABLE

I	Summary of Null Distribution of $S^+(p) =$ $\max_{1 \leq k \leq Np} S_k^* / \{\text{var}(S_{Np}^* H_0)\}^{1/2}$ in the case $m = n = N/2$	15
II	Summary of Null Distribution of $S(p) =$ $\max_{1 \leq k \leq Np} S_k^* / \{\text{var}(S_{Np}^* H_0)\}^{1/2}$ in the case $m = n = N/2$	17
III	Simulation Estimates of $\text{pr}\{S^+(p) \geq x_\infty^\alpha\}$	20
IV	Simulation Estimates of $\text{pr}\{S(p) \geq x_\infty^\alpha\}$	21
V	Fitted Critical Values for KP Procedures	24
VI	Simulation Estimates of Actual Levels Corresponding to Fitted Critical Values for KP Procedures	28
VII	Simulation Estimates of Actual Levels Corresponding to One-Sided Fitted Critical Values in the Unbalanced Case, with $N = 100$, $\alpha = .05$	31
VIII	Simulation Estimates of Actual Levels Corresponding to Two-Sided Fitted Critical Values in the Unbalanced Case, with $N = 100$, $\alpha = .05$	32
IX	Simulation Results for Operating Characteristics of the One- Sided Tests Under Lehmann Alternatives, with $m = n = N/2$, $\alpha = .05$..	36
X	Simulation Results for Operating Characteristics of the Two- Sided Tests Under Lehmann Alternatives, with $m = n = N/2$, $\alpha = .05$..	39
XI	Expected Proportions Sampled Using Early Decision Rules in the One-Sided Case, with $\alpha = .05$, $p = 1$	48
XII	Expected Proportions Sampled Using Early Decision Rules in the Two-Sided Case, with $\alpha = .05$, $p = 1$	49
XIII	Statistics for the First Example	51
XIV	Mouse Survival Data	53

LIST OF FIGURES

FIGURE

- 1 Estimated critical values for the one-sided sequential
 KP procedure at nominal levels .10, .05 and .01 25
- 2 Estimated critical values for the two-sided sequential
 KP procedure at nominal levels .10, .05 and .01 26

ACKNOWLEDGEMENT

I would like to thank Professor A. John Petkau for his guidance and assistance in the producing of this thesis. I also appreciate the assistance of Professor James A. Koziol. I would like to thank Janet Clark and Mary Daisley for the typing of this thesis. Finally, I would like to thank the National Research Council of Canada for supporting me with an NRC Postgraduate Scholarship.

Chapter 1

INTRODUCTION

A common situation in life testing is assessing whether two treatments are equally effective in prolonging life. We suppose that N subjects are to be treated (m subjects with the first treatment, n subjects with the second treatment, $N = m + n$) and that all are available at the same time to begin experimentation. The observations are lengths of survival or times until failure and are naturally ordered. We assume that the failure times from the two treatments have respective continuous distribution functions (survival distributions) F_1 and F_2 . We then want to test the null hypothesis H_0 that the survival distributions are equal (i.e. $F_1 = F_2$) against either the alternative hypothesis H_1 that failure times from the first treatment are stochastically larger (i.e. $F_1 < F_2$, the first treatment is superior) or the alternative hypothesis H_2 that the survival distributions are unequal (i.e. $F_1 \neq F_2$). Many nonparametric tests are available for the situation where all failure times have been observed, the Savage and Wilcoxon rank tests being perhaps most commonly used. However in some life tests, such as clinical trials, it is desirable to make a decision concerning the relative efficacies of competing treatments before all the failure times are known. Two different types of procedures that permit early decisions are the fixed point censoring procedure and the sequential or progressive censoring procedure.

In a fixed point censoring procedure, a decision is made after observing, say, the first r failure times (where r is prespecified). Gastwirth (1965) and Johnson and Mehrotra (1972) have indicated how the Savage, Wilcoxon and other linear rank tests can be appropriately modified for this fixed point

censoring situation. In a sequential or progressive censoring procedure, at most r failure times are observed and a decision is made whether to continue or terminate the experiment after observing each failure time. Thus, the experiment can be terminated earlier if evidence accumulates against the null hypothesis. Chatterjee and Sen (1973) investigated a general class of sequential procedures based on linear rank statistics and, among other results, determined asymptotic critical values for these procedures. Subsequently, specific sequential procedures have been investigated by Muenz, Green and Byar (1977), Davis (1978) and Koziol and Petkau (1978); these procedures are respectively based on the Mantel-Haenszel, Wilcoxon and Savage statistics. The theory of Chatterjee and Sen provided asymptotic critical values for the second and third of these procedures.

In life testing, exponential, and more generally Weibull, distributions are often used to model empirical survival distributions; therefore any non-parametric procedure that is to be used in practice should perform well when the underlying distribution functions F_1 and F_2 are exponential or Weibull. Thus, different competing procedures are often compared on the basis of power under Lehmann alternatives (subclasses of which are the exponential and Weibull scale shift alternatives). Sequential procedures are also compared on the basis of expected proportion sampled, the average proportion of the N subjects observed before a decision is made; for a fixed point censoring procedure with censoring point r , the expected proportion sampled is r/N . Using a procedure with low values of expected proportion sampled will likely reduce the cost of an experiment but there could be an increased risk of an incorrect decision. Both the power and expected proportion sampled should therefore be considered in choosing a procedure.

Muenz, Green and Byar (1977, hereafter referred to as MGB) presented

no asymptotic results relevant to their sequential procedure based on the Mantel-Haenszel statistic. Instead, they numerically examined the null distributions of their proposed statistic and the usual Mantel-Haenszel statistic and determined approximate critical values for these statistics in the special case of $m=n=N/2$, $r=N$. Also, they indicated how the possibility of early acceptance of the null hypothesis could be incorporated into their sequential procedure and the corresponding fixed point censoring procedure. They numerically compared the two procedures on the basis of power under Lehmann alternatives and expected proportion sampled (in the case $m=n=N/2$, $r=N$). For the sequential procedure based on the Wilcoxon statistic, Davis (1978) examined the accuracy of the asymptotic critical values (provided by the theory of Chatterjee and Sen) for finite samples. He numerically compared this procedure with the fixed point censoring procedure due to Halperin and Ware (1974) for the case of exponential scale shift alternatives. The latter procedure has a curtailed form which incorporates the possibility of early acceptance of the null hypothesis. For the sequential procedure based on the Savage statistic, Koziol and Petkau (1978, hereafter referred to as KP) presented asymptotic distributions of their proposed test statistics in the case of contiguous exponential scale shift alternatives (they later discovered that similar asymptotic results hold in the more general case of contiguous Weibull scale shift alternatives). In addition, they compared the asymptotic properties, power and expected proportion sampled under contiguous exponential scale shift alternatives, of their sequential procedure with those of the corresponding fixed point censoring procedure. The numerical studies of MGB and Davis and the theoretical investigation of KP demonstrated that the differences in power between the sequential procedures and their fixed point censoring counterparts were quite small for larger sample sizes; moreover, the sequential procedures offer

considerable reduction in the expected proportion sampled. Although these results establish the attractiveness of the sequential procedures, no detailed comparisons among the different sequential procedures have yet been carried out. On the basis of the limited numerical results presented by both MGB and Davis, one cannot definitively compare their competing sequential procedures. Also, comparison of the asymptotic results of KP to the Monte Carlo results of MGB and Davis seems inappropriate. Considering the nature of the stopping rules of the sequential procedures based on the Savage and Mantel-Haenszel statistics (see Section 6.1 for further explanation), KP conjectured that the sequential procedure based on the Savage statistic will generally be more powerful (under Lehmann alternatives) than that based on the Mantel-Haenszel statistic, while the latter procedure will generally be quicker to come to a decision. Also, since the Savage rank test is the locally most powerful rank test under Lehmann alternatives (Savage, 1956) and the sequential procedure proposed by Davis is identical to that proposed by KP except Wilcoxon scores are used instead of Savage scores, one might expect that the latter procedure will be more powerful than the former.

The purposes of this thesis are to determine the sample size required before the asymptotic critical values for the sequential procedure proposed by KP provide a reasonable approximation, and more importantly, to compare in detail the sequential procedures proposed by MGB, Davis and KP. We formally describe all the procedures under consideration in Chapter 2. In Chapter 3, we examine the accuracy of the asymptotic critical values for finite samples for the sequential procedure based on the Savage statistic. There we indicate how we numerically determined approximate critical values for the sequential procedures. In Chapter 4, we numerically compare the sequential procedures to each other and to their fixed point censoring counterparts. We indicate in Chapter 5 how early decision rules, which

include the possibility of early acceptance of the null hypothesis, can be incorporated into the procedures based on the Savage and Wilcoxon statistics. We provide some indication of the reduction in the expected proportion sampled which can be realized with their implementation. We present two examples in Chapter 6, and discuss our results in Chapter 7.

Chapter 2

DESCRIPTION OF PROCEDURES

Section 2.1 INTRODUCTION

Let $X_1, \dots, X_m$ and $X_{m+1}, \dots, X_N$ be independent random samples of sizes m and n ($n = N - m$) from populations π_1 and π_2 with respective continuous distribution functions F_1 and F_2 . We are interested in testing the null hypothesis $H_0 : F_1 = F_2$ against either the one-sided alternative $H_1 : F_1 < F_2$ or the two-sided alternative $H_2 : F_1 \neq F_2$. Let $X = (X_{(1)}, \dots, X_{(N)})$ denote the vector of order statistics, and let $D = (d_1, \dots, d_N)$ denote the vector of indicator variables, where $d_i = 1$ if the i th order statistic, $X_{(i)}$, of the combined sample is from population π_1 , $d_i = 0$ if it is from π_2 . In Sections 2.2, 2.3 and 2.4, the fixed point and progressive censoring procedures based on the Savage, Wilcoxon and Mantel-Haenszel statistics are described. In Section 2.5, a modification of the sequential procedure proposed by MGB is outlined.

Section 2.2 PROCEDURES BASED ON THE SAVAGE STATISTIC

When all observations are available, the Savage statistic can be written as (cf. Savage, 1956; Hájek and Šidák, 1967, p. 97)

$$S_N = \sum_{i=1}^N d_i a_N(i)$$

where $a_N(i) = b_N(i) - 1$ and the scores $b_N(i)$ are defined by $b_N(i) = \sum_{j=1}^i (N-j+1)^{-1}$. It is well-known that $E(S_N | H_0) = 0$ and

$$\text{var}(S_N | H_0) = \frac{mn}{N-1} (1 - b_N(N)/N).$$

If only the first k of the N order statistics are to be observed, the appropriate modification of the Savage statistic can be written as (cf. Gastwirth, 1965; Johnson and Mehrotra, 1972)

$$S_k^* = \sum_{i=1}^N d_i a_N^k(i)$$

where $a_N^k(i) = b_N^k(i) - 1$ and

$$b_N^k(i) = b_N(i) \quad i = 1, \dots, k$$

$$= \sum_{j=k+1}^N b_N(j) / (N-k) \quad i = k+1, \dots, N.$$

Note that $S_{N-1}^* = S_N^* = S_N$. It is easily verified that $E(S_k^* | H_0) = 0$ and

$$\text{var}(S_k^* | H_0) = \frac{mn}{N-1} \left(\frac{k}{N} - b_N(k)/N \right)$$

The fixed point censoring procedure with censoring point $r = Np$ ($0 < p \leq 1$) is to reject H_0 for the one-sided alternative if $S_r^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ is too large, or for the two-sided alternative if $|S_r^*| / \{\text{var}(S_r^* | H_0)\}^{1/2}$ is too large. Since under H_0 $|S_r^*| / \{\text{var}(S_r^* | H_0)\}^{1/2}$ is asymptotically (as $\min(m, n) \rightarrow \infty$) distributed as a standard normal random variable, the asymptotic critical values at the α -level of significance are $z_{1-\alpha}$ and $z_{1-\alpha/2}$ respectively for the one-sided and two-sided tests. (Note: z_β satisfies $\Phi(z_\beta) = \beta$, where Φ is the standard normal cumulative distribution function).

The progressive censoring procedure can be described as follows. For the one-sided alternative, progressively determine for every k , $1 \leq k \leq r$, whether $S_k^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ exceeds some critical value x ; if so terminate observation and reject H_0 , if not, continue. If $S_k^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$, $1 \leq k \leq r$, are all less than x , then H_0 can be accepted (i.e. H_0 is rejected

if and only if $\max_{1 \leq k \leq r} S_k^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ is too large). For the two-sided

alternative, a similar procedure based on $|S_k^*|$ is appropriate. It has been shown by Chatterjee and Sen that as $\min(m, n) \rightarrow \infty$ and $r/N \rightarrow p$ ($0 < p \leq 1$),

$$\lim_{N \rightarrow \infty} \text{pr} \left\{ \max_{1 \leq k \leq r} S_k^* / [\text{var}(S_r^* | H_0)]^{1/2} \geq x \right\} = 2[1 - \Phi(x)] \quad (2.1)$$

$$\lim_{N \rightarrow \infty} \text{pr} \left\{ \max_{1 \leq k \leq r} |S_k^*| / [\text{var}(S_r^* | H_0)]^{1/2} \geq x \right\} = 4 \sum_{k=0}^{\infty} (-1)^k [1 - \Phi([2k+1]x)] \quad (2.2)$$

for every $x > 0$. For the larger values of x , the right hand side of (2.2) can be accurately approximated by $4[1 - \Phi(x)]$ (see KP, 1978, Table 1). The asymptotic critical value at the α -level of significance is $z_{1-\alpha/2}$ for the one-sided test; for the two-sided test, it is approximately $z_{1-\alpha/4}$ for α less than .30.

Section 2.3 PROCEDURES BASED ON THE WILCOXON STATISTIC

The Wilcoxon statistic can be written as (cf. Wilcoxon, 1945; Hájek and Šidák, 1967, p. 87)

$$W_N = \sum_{i=1}^N d_i a_N(i)$$

where $a_N(i) = b_N(i) - (N+1)/2$ and the scores have the form $b_N(i) = i$. It is well-known that $E(W_N | H_0) = 0$ and $\text{var}(W_N | H_0) = mn(N+1)/12$. If only the first k order statistics are to be observed, the appropriate modification of the Wilcoxon statistic can be written as

$$W_k^* = \sum_{i=1}^N d_i a_N^k(i)$$

where $a_N^k(i) = b_N^k(i) - (N+1)/2$ and

$$b_N^k(i) = b_N(i) \quad i = 1, \dots, k$$

$$= \sum_{j=k+1}^N b_N(j) / (N-k) \quad i = k+1, \dots, N.$$

It is easily verified that $E(W_k^* | H_0) = 0$ and

$$\text{var } (W_k^* | H_0) = \frac{mn(N+1)}{12} [1 - \{N^2(1-p)^3 - (1-p)\} / (N^2-1)]$$

where $p = k/N$. In both the fixed point and progressive censoring situations, the procedures based on the Wilcoxon statistics are completely analogous to those based on the Savage statistics and in each case the asymptotic critical values are identical.

Remark. The fixed point censoring procedure based on the Wilcoxon statistic which is being considered here is different from the fixed point censoring procedure due to Halperin and Ware (1974). Their procedure prescribes termination of the experiment as soon as a proportion p of the failure times has been observed in one of the two treatment groups; this procedure would be somewhat quicker to terminate, particularly when the null hypothesis is false.

Section 2.4 PROCEDURES BASED ON THE MANTEL-HAENSZEL STATISTIC

The Mantel-Haenszel statistic can be motivated as follows. After the i th failure time in the combined sample, for the 2×2 table

	group I	group II	total
failed	d_i	$1 - d_i$	1
survived	$m - \sum_{j=1}^i d_j$	$n - i + \sum_{j=1}^i d_j$	$N - i$
total	$m - \sum_{j=1}^{i-1} d_j$	$n - i + 1 + \sum_{j=1}^{i-1} d_j$	$N - i + 1$

(by convention, $\sum_{j=1}^0 d_j = 0$). Under the null hypothesis, the expected value of D_i given $(d_1, \dots, d_{i-1})$ is $e_i = (m - \sum_{j=1}^{i-1} d_j) / (N - i + 1)$ and the variance of D_i given $(d_1, \dots, d_{i-1})$ is $v_i = e_i(1 - e_i)$; the quantities e_i and v_i are suggested from the hypergeometric distribution associated with the 2×2 table. The Mantel-Haenszel statistic (cf. Mantel and Haenszel, 1959; Mantel, 1966), when all the observations are available, is

$$X_N = [|\sum_{i=1}^N (d_i - e_i)| - \frac{1}{2}]^2 / \sum_{i=1}^N v_i.$$

Under H_0 this statistic is asymptotically chi-squared with one degree of freedom. If only the first k order statistics are to be observed, the modification of the Mantel-Haenszel statistic is

$$X_k^* = [|\sum_{i=1}^k (d_i - e_i)| - \frac{1}{2}]^2 / \sum_{i=1}^k v_i.$$

MGB consider the Mantel-Haenszel statistic without the continuity correction; that is, they use

$$M_k^{*2} = [\sum_{i=1}^k (d_i - e_i)]^2 / \sum_{i=1}^k v_i.$$

We use the unsquared version of this statistic,

$$M_k^* = \sum_{i=1}^k (e_i - d_i) / \left\{ \sum_{i=1}^k v_i \right\}^{1/2} = S_k^* / \left\{ \sum_{i=1}^k v_i \right\}^{1/2}. \quad (2.3)$$

The last equality can be easily verified.

The fixed point censoring procedure with censoring point r is to reject H_0 for the one-sided alternative if M_r^* (or $\text{sign} \left\{ \sum_{i=1}^r (e_i - d_i) \right\} M_r^{*2}$, which is of the form of the one-sided statistic MGB use) is too large, or for the two-sided alternative if $|M_r^*|$ is too large. KP have shown $\sum_{i=1}^r v_i$ to be a consistent (as $N \rightarrow \infty$ and $m/N \rightarrow \lambda$, $0 < \lambda < 1$) estimator of $\text{var}(S_r^* | H_0)$, therefore the fixed point censoring procedures based on the Mantel-Haenszel and Savage statistics are asymptotically equivalent.

The progressive censoring procedure with at most r order statistics to be observed is to reject H_0 for the one-sided alternative if $\max_{1 \leq k \leq r} M_k^*$ is too large, or for the two-sided alternative if $\max_{1 \leq k \leq r} |M_k^*|$ is too large (cf. Section 2.2). MGB did not provide asymptotic critical values for these tests, but instead provided critical value curves (obtained by smoothing Monte Carlo critical values) in the special case of equal sample sizes ($m = n = N/2$) with $r = N$.

Section 2.5 THE MODIFIED MGB SEQUENTIAL PROCEDURE

We consider a slight modification of the procedures based on the Mantel-Haenszel statistic. This modification, M_k' , consists of replacing

$\sum_{i=1}^k v_i$ in the definition of M_k^* , as given in (2.3), by $\text{var}(S_k^* | H_0)$ (i.e. $M_k' = S_k^* / \{\text{var}(S_k^* | H_0)\}^{1/2}$). The test statistics for the sequential procedure are

now $\max_{\underline{l} < \underline{k} < \underline{r}} M'_k$ and $\max_{\underline{l} < \underline{k} < \underline{r}} |M'_k|$ for the one-sided and two-sided tests

respectively. Since $\sum_{i=1}^k v_i$ is a consistent estimator of $\text{var}(S_k^* | H_0)$, this

modification has no effect asymptotically. In the fixed point censoring case, the modified procedure is identical to the procedure based on the Savage statistic.

Chapter 3

NULL DISTRIBUTION RESULTS

Section 3.1 INTRODUCTION

We will now refer to the sequential procedures as the KP, Davis, MGB and modified MGB procedures.

In this chapter, one objective is to examine the accuracy of the asymptotic critical values for the KP procedure for finite samples. It was found that use of the asymptotic critical values for the KP procedure resulted in somewhat conservative tests, even for moderate-sized samples. Therefore, in the equal sample size case, empirical critical values were simulated and fitted by smooth curves. The fitted critical values and curves are presented in Section 3.2; also presented are exact critical values in the equal sample size cases $N = 10, 12, 14, 16, 18, 20$ with $p = .5$ and 1 . Similar methods were used to determine approximate critical values in the equal sample size case for the Davis, MGB and modified MGB procedures; this is discussed in Section 3.3. In Section 3.4, we study the effect of unequal sample sizes ($m \neq n$) on the critical values for the four sequential procedures. In Section 3.5, we examine the accuracy of the asymptotic critical values for the fixed point censoring procedures based on the Savage, Wilcoxon and Mantel-Haenszel statistics.

Section 3.2 THE KP PROCEDURE, WITH EQUAL SAMPLE SIZES

We consider the case in which at most the first $r = Np$ ($0 < p \leq 1$) of the N order statistics are to be observed and examine the null distributions of the following statistics:

$$S^+(p) = \max_{1 \leq k \leq Np} S_k^* / \{\text{var}(S_{Np}^* | H_0)\}^{1/2}$$

$$S^-(p) = \min_{1 \leq k \leq Np} S_k^* / \{\text{var}(S_{Np}^* | H_0)\}^{1/2}$$

$$S(p) = \max_{1 \leq k \leq Np} |S_k^*| / \{\text{var}(S_{Np}^* | H_0)\}^{1/2} .$$

For small values of Np , the null distributions of these statistics can be determined by complete enumeration of the D -vectors. We have done this in the case of equal sample sizes, $m=n=N/2$, for $N = 10(2)20$ and $p = .5, 1$. Some results for the null distribution of $S^+(p)$ are summarized in Table I. For each N and each nominal level α , $\text{pr}\{S^+(p) \geq x\}$ is tabulated for those values of x yielding levels equal to or closest to the nominal level. Also tabulated are the actual significance levels corresponding to the asymptotic critical values. The corresponding results for the null distribution of $S(p)$ are summarized in Table II.

When examining the detailed listings of these null distributions, the equality

$$\text{pr}\{S(p) \geq x\} = 2 \text{pr}\{S^+(p) \geq x\}$$

was found to hold for every value of x greater than 1.20. For the case of equal sample sizes, we have

$$\text{pr}\{S^+(p) \geq x\} = \text{pr}\{S^-(p) \leq -x\} \quad (3.1)$$

and consequently

$$\text{pr}\{S(p) \geq x\} = 2\text{pr}\{S^+(p) \geq x\} - \text{pr}\{S^+(p) \geq x, S^-(p) \leq -x\}. \quad (3.2)$$

Since one would expect the second term on the right of (3.2) to be small

Table I. Summary of Null Distribution of $S^+(p) = \max_{1 \leq k \leq N_p} S_k^* / \{\text{var}(S_{N_p}^* | H_0)\}^{1/2}$ in the case $m = n = N/2$

A. Case $p = 1/2$

Nominal Level α		.01		.05		.10		.20	
Asymp. crit. value x_∞^α		2.5758		1.9600		1.6449		1.2816	
N = 10	x	2.93525	2.17754	1.87445	1.74455	1.63090	1.52987	1.13838	1.02472
	P(x)	.00397	.02381	.04365	.06349	.08333	.10317	.18254	.22222
	$P(x_\infty^\alpha)$	.00397		.02381		.06349		.14286	
N = 12	x	2.53578	2.29918	1.94556	1.91469	1.51214	1.42933	1.07443	1.05372
	P(x)	.00758	.01407	.04004	.05628	.08874	.10498	.19156	.20779
	$P(x_\infty^\alpha)$	.00108		.03355		.05628		.15909	
N = 14	x	2.50633	2.43675	1.79897	1.74010	1.49701	1.45310	1.28083	1.24120
	P(x)	.00845	.01049	.04516	.05128	.09819	.10431	.17162	.20221
	$P(x_\infty^\alpha)$	.00641		.02681		.07372		.16550	
N = 16	x	2.36928	2.30429	1.82886	1.81493	1.59055	1.58721	1.17269	1.14726
	P(x)	.00940	.01158	.04856	.05074	.08990	.10404	.18780	.21717
	$P(x_\infty^\alpha)$	.00723		.03333		.07902		.14646	
N = 18	x	2.38616	2.37675	1.80557	1.79755	1.48015	1.47812	1.17599	1.16673
	P(x)	.00983	.01057	.04994	.05167	.09831	.10004	.19852	.20025
	$P(x_\infty^\alpha)$	.00465		.03723		.06981		.14842	
N = 20	x	2.45476	2.43631	1.80666	1.80114	1.53333	1.52844	1.11532	1.10848
	P(x)	.00988	.01013	.04950	.05015	.09926	.10289	.19856	.20107
	$P(x_\infty^\alpha)$	.00631		.03197		.07701		.15927	

Table I. (continued)

B. Case $p = 1$

Nominal Level α		.01		.05		.10		.20	
Asymp. crit. value x_{∞}^{α}		2.5758		1.9600		1.6449		1.2816	
N = 10	x	2.16068	2.04176	1.72068	1.70878	1.43442	1.41063	1.17279	1.14023
	P(x)	.00794	.01190	.04365	.05159	.09921	.10317	.19841	.20238
	$P(x_{\infty}^{\alpha})$	0		.01587		.05556		.13492	
N = 12	x	2.10872	2.10159	1.70214	1.69424	1.47539	1.46955	1.11111	1.10933
	P(x)	.00974	.01082	.04978	.05087	.09957	.10065	.19805	.20022
	$P(x_{\infty}^{\alpha})$	0		.02056		.05411		.13853	
N = 14	x	2.15557	2.14459	1.73078	1.72705	1.44734	1.44708	1.11729	1.11612
	P(x)	.00991	.01020	.04983	.05012	.09994	.10023	.19843	.20047
	$P(x_{\infty}^{\alpha})$	.00058		.02127		.05944		.15181	
N = 16	x	2.17564	1.17368	1.71965	1.71940	1.46137		1.14658	1.14577
	P(x)	.00995	.01002	.04996	.05004	.10		.19992	.20008
	$P(x_{\infty}^{\alpha})$	.00101		.02378		.06208		.15058	
N = 18	x	2.19731	2.19719	1.73829	1.73822	1.47292		1.14808	1.14804
	P(x)	.009996	.01002	.04994	.05016	.10		.19914	.20134
	$P(x_{\infty}^{\alpha})$	.00144		.02489		.06399		.15121	
N = 20	x	2.22034	2.22033	1.74874	1.74872	1.47955	1.47947	1.15242	1.15238
	P(x)	.009997	.010002	.049996	.050001	.099997	.100008	.19999	.20003
	$P(x_{\infty}^{\alpha})$	.00178		.02621		.06716		.15533	

Note: $P(x) = \text{pr}\{S^+(p) \geq x\}$.

Table II. Summary of Null Distribution of $S(p) = \max_{1 \leq k \leq Np} |S_k^*| / \{\text{var}(S_{Np}^* | H_0)\}^{1/2}$ in the case $m = n = N/2$

A. Case $p = 1/2$

Nominal Level		.01		.05		.10		.20	
Asymp. crit. value		2.8070		2.2414		1.9600		1.6449	
x_∞^α									
N = 10	x	2.93525	2.17754	2.17754	1.87445	1.87445	1.74455	1.63090	1.52987
	P(x)	.00794	.04762	.04762	.08730	.08730	.12698	.16667	.20635
	$P(x_\infty^\alpha)$	.00794		.00794		.04762		.12698	
N = 12	x	3.24559	2.53578	2.19567	2.10365	1.94556	1.91469	1.51214	1.42933
	P(x)	.00216	.01515	.04113	.05411	.08009	.11255	.17749	.20996
	$P(x_\infty^\alpha)$	.00216		.02814		.06710		.11255	
N = 14	x	2.66789	2.58285	2.26397	2.00885	1.79897	1.74010	1.49071	1.45310
	P(x)	.00874	.01282	.04138	.05361	.09033	.10256	.19639	.20862
	$P(x_\infty^\alpha)$	.00466		.04138		.05361		.14744	
N = 16	x	2.69470	2.64703	2.06598	2.06382	1.82886	1.81493	1.59055	1.58721
	P(x)	.00886	.01010	.04926	.05796	.09713	.10148	.17980	.20808
	$P(x_\infty^\alpha)$	.00513		.02751		.06667		.15804	
N = 18	x	2.58815	2.54005	2.13379	2.12408	1.80557	1.79755	1.48015	1.47812
	P(x)	.00930	.01078	.04879	.05027	.09988	.10333	.19663	.20008
	$P(x_\infty^\alpha)$	.00485		.03348		.07445		.13961	
N = 20	x	2.66495	2.64515	2.05488	2.04124	1.80666	1.80114	1.53333	1.52844
	P(x)	.00970	.01019	.04997	.05127	.09901	.10031	.19851	.20578
	$P(x_\infty^\alpha)$	.00450		.03292		.06393		.15402	

Table II. (continued)

B. Case $p = 1$

Nominal Level α		.01		.05		.10		.20	
Asymp. crit. value x_{∞}^{α}		2.8070		2.2414		1.9600		1.6449	
N = 10	x	2.30338	2.16068	1.86338	1.85064	1.72068	1.70878	1.43442	1.41063
	$P(x)$	.00794	.01587	.04762	.05556	.08730	.10317	.19841	.20636
	$P(x_{\infty}^{\alpha})$	0		.00794		.03175		.11111	
N = 12	x	2.28068	2.23712	1.91484	1.91002	1.70214	1.69424	1.47539	1.46955
	$P(x)$	.00866	.01082	.04978	.05195	.09957	.10173	.19913	.20130
	$P(x_{\infty}^{\alpha})$	0		.00866		.04113		.10823	
N = 14	x	2.27755	2.27221	1.91882	1.91064	1.73078	1.72705	1.44734	1.44708
	$P(x)$	.00991	.01049	.04953	.05012	.09965	.10023	.19988	.20047
	$P(x_{\infty}^{\alpha})$	0		.01224		.04254		.11888	
N = 16	x	2.31515	2.31212	1.94290	1.94095	1.71965	1.71940	1.46137	
	$P(x)$	.00995	.01010	.04988	.05004	.09992	.10008	.20	
	$P(x_{\infty}^{\alpha})$	.00031		.01430		.04755		.12416	
N = 18	x	2.34872	2.34758	1.95843	1.95775	1.73829	1.73822	1.47292	
	$P(x)$	.009996	.01004	.04998	.05002	.09988	.10033	.20	
	$P(x_{\infty}^{\alpha})$	.00049		.01645		.04977		.12797	
N = 20	x	2.37883	2.37850	1.97456	1.97427	1.74874	1.74872	1.47955	1.47947
	$P(x)$	.00999	.010002	.04998	.050001	.09999	.100002	.19999	.20002
	$P(x_{\infty}^{\alpha})$	.00076		.01823		.05242		.13433	

Note: $P(x) = \text{pr} \{ S(p) \geq x \}$.

for larger values of x , the above empirical result is not unexpected. The results in Tables I and II clearly indicate that, at least for these sample sizes, the use of the asymptotic critical values for the KP procedure will result in tests which are unduly conservative.

Although the cost of determining the null distributions exactly becomes prohibitive as Np increases, the distributions can still be easily simulated by random sampling of the D-vectors. A simulation experiment was carried out in order to determine how large the sample size must be before the asymptotic critical values provide a reasonable approximation. Table III summarizes the results for the one-sided test at nominal levels .10, .05 and .01. Each entry in the table is based on 1500 simulated samples; entries for a fixed N but different values of p and α were obtained from the same set of simulated samples. In calculating the simulation estimates of $\text{pr}\{S^+(p) \geq x\}$, (3.1) was used to form two different estimates which were then averaged to form the final estimate; the standard error of any individual entry $\hat{p}$ in Table III is estimated to be at most $\{\hat{p}(1-\hat{p})/3000\}^{1/2}$. The corresponding results for the two-sided case are presented in Table IV (the same set of simulated samples as in the one-sided case was used).

The results in Tables III and IV clearly indicate that, even for what would in most instances be thought of as large sample sizes, the use of the asymptotic critical values for the KP procedure will result in tests which are somewhat conservative. Note also the tendency of the tests to be more conservative for the case $p = 1.0$ than for the cases $p = .50, .75$ and $.90$. These results are in contrast to the corresponding results provided by Davis (1978; Table III, p. 394) who found that for the one-sided Davis procedure, the asymptotic critical value (at nominal level .05) provides an adequate approximation (although still yielding a slightly conservative test) for

Table III. Simulation Estimates of $\text{pr} \{S^+(p) \geq x_\infty^\alpha\}^*$.

m=n	P				
	.25	.50	.75	.90	1.0
$\alpha = .10$					
10	.039	.065	.068	.063	.059
20	.085	.079	.085	.084	.075
30	.080	.083	.088	.087	.085
40	.084	.084	.083	.088	.081
50	.087	.092	.097	.094	.088
70	.090	.096	.093	.089	.085
100	.079	.085	.086	.087	.085
200	.090	.093	.099	.099	.099
$\alpha = .05$					
10	.015	.027	.031	.026	.022
20	.042	.040	.037	.036	.033
30	.045	.041	.042	.041	.035
40	.038	.041	.039	.041	.039
50	.047	.048	.047	.046	.044
70	.044	.049	.044	.044	.042
100	.038	.044	.042	.041	.036
200	.043	.047	.041	.051	.051
$\alpha = .01$					
10	.000	.005	.004	.002	.001
20	.006	.006	.005	.005	.004
30	.006	.007	.010	.008	.007
40	.005	.008	.008	.008	.006
50	.008	.011	.009	.007	.006
70	.008	.010	.008	.006	.006
100	.009	.010	.007	.005	.006
200	.006	.012	.011	.011	.009

*Each entry is based on 1500 simulated samples.

Table IV. Simulation Estimates of $\text{pr}\{S(p) \geq x_{\infty}^{\alpha}\}^*$.

m=n	P				
	.25	.50	.75	.90	1.0
$\alpha = .10$					
10	.030	.054	.061	.052	.045
20	.084	.079	.074	.073	.066
30	.089	.083	.083	.083	.069
40	.077	.083	.079	.083	.078
50	.095	.096	.093	.091	.089
70	.088	.099	.088	.087	.083
100	.077	.089	.083	.082	.072
200	.085	.095	.103	.102	.102
$\alpha = .05$					
10	.030	.029	.033	.021	.019
20	.025	.040	.033	.029	.025
30	.034	.035	.042	.037	.033
40	.034	.044	.039	.039	.040
50	.041	.049	.049	.044	.037
70	.043	.050	.043	.044	.038
100	.036	.045	.041	.035	.031
200	.042	.059	.056	.057	.053
$\alpha = .01$					
10	.000	.003	.002	.000	.000
20	.010	.007	.007	.004	.003
30	.003	.008	.009	.008	.007
40	.003	.007	.007	.009	.005
50	.009	.013	.009	.005	.004
70	.009	.010	.005	.007	.003
100	.009	.010	.006	.007	.005
200	.005	.012	.014	.011	.009

*Each entry is based on 1500 simulated samples.

situations in which $Np > 60$.

That the asymptotic results are much slower to take hold with the KP procedure than with the Davis procedure might be explained upon evaluation of the moments of the modified Savage and Wilcoxon statistics. Letting $SS_r = S_r^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ and $SW_r = W_r^* / \{\text{var}(W_r^* | H_0)\}^{1/2}$, calculations (Hájek and Šidák, 1967, p. 81-82) in the equal sample size case yield

$$E(SS_r^3) = E(SW_r^3) = 0$$

$$E(SS_r^4) = 3 - f_s\left(\frac{r}{N}, \frac{1}{2}\right) / N + O(N^{-2})$$

$$E(SW_r^4) = 3 - f_w\left(\frac{r}{N}, \frac{1}{2}\right) / N + O(N^{-2})$$

where

$$f_s(p, \lambda) = 6 + p^{-2} [6 - \lambda^{-1} (1 - \lambda)^{-1}] [-3p^2 + 9p - 6(1 - p) \ln^2(1 - p) + 8(1 - p) \ln(1 - p)],$$

$$f_w(p, \lambda) = 6 + .6p [6 - \lambda^{-1} (1 - \lambda)^{-1}] [1 - (1 - p)^3]^{-2} [15 - 105p + 210p^2 - 180p^3 + 63p^4 - 5p^5].$$

$f_s(p, \frac{1}{2})$ evaluated at $p = .25, .50, .75, .90, 1.0$ is respectively 4.847, 2.288, 3.892, 7.597, 18.000. Similarly, $f_w(p, \frac{1}{2})$ evaluated at $p = .25, .50, .75, .90, 1.0$ is respectively 5.375, 3.086, 3.385, 3.578, 3.600. The usual (Edgeworth) correction to the asymptotic distribution functions of these statistics then become (Hájek and Šidák, 1967, p. 146; see also Bickel and van Zwet, 1978; Robinson, 1978)

$$\text{pr}\{SS_r \leq x\} \cong \Phi(x) - \phi^{(4)}(x) f_s\left(\frac{r}{N}, \frac{1}{2}\right) / 24N \quad (3.3)$$

$$\text{pr}\{SW_r \leq x\} \cong \Phi(x) - \phi^{(4)}(x) f_w\left(\frac{r}{N}, \frac{1}{2}\right) / 24N \quad (3.4)$$

For the larger values of $p = r/N$, because $f_w(p, \frac{1}{2}) \leq f_s(p, \frac{1}{2})$, (3.3) and (3.4) indicate that the asymptotic distribution provides a better approximation for

SW_r than for SS_r . One might then conjecture for these values of $p = r/N$ that the asymptotic distributions (2.1), (2.2) provide better approximations for the Davis statistics, $\max_{1 \leq k \leq r} W_k^* / \{\text{var}(W_r^* | H_0)\}^{1/2}$ and $\max_{1 \leq k \leq r} |W_k^*| / \{\text{var}(W_r^* | H_0)\}^{1/2}$ than for the corresponding KP statistics, $\max_{1 \leq k \leq r} S_k^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ and $\max_{1 \leq k \leq r} |S_k^*| / \{\text{var}(S_r^* | H_0)\}^{1/2}$.

Since the asymptotic critical values are of little use for the sample sizes of usual interest, the set of simulated samples used to produce Tables III and IV was also used to determine empirical critical values for the KP procedure. These empirical critical values for the one-sided test are plotted in Figure 1 for the nominal levels .10, .05 and .01. Critical values for a given value of N but different values of p and α were obtained from the same set of 1500 simulated samples. The "exact" critical values (values yielding levels closest to the nominal level) for the case $N = 20$ are also plotted on this figure. Similarly, the analogous results for the two-sided test are plotted in Figure 2.

To remove some of the sampling variability evidenced in the figures, smooth curves were fit to the empirical critical values. Guided by the asymptotic theory, various simple curves depending on N but not on p were tested. For the one-sided case, curves of the form $S_{N,\alpha}^+ = a_\alpha - b_\alpha \exp(-c_\alpha N^{1/2})$ where a_α is the asymptotic critical value appeared to perform well and were fitted by non-linear least squares. Similarly for the two-sided case, curves of the form $S_{N,\alpha} = A_\alpha - B_\alpha \exp(-C_\alpha N^{1/2})$ where A_α is the asymptotic critical value were fitted. The values of the fitted coefficients and the resulting "smoothed" empirical critical values are provided in Table V, and the curves are depicted in Figures 1 and 2.

Table V. Fitted Critical Values for KP Procedures*

m=n	$S_{N,\alpha}^+$			$S_{N,\alpha}$		
	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$	$\alpha = .10$	$\alpha = .05$	$\alpha = .01$
10	1.5113	1.7956	2.3296	1.7956	2.0360	2.5432
15	1.5268	1.8194	2.3719	1.8194	2.0692	2.5907
20	1.5384	1.8368	2.4018	1.8368	2.0930	2.6241
25	1.5477	1.8504	2.4245	1.8504	2.1112	2.6492
30	1.5554	1.8613	2.4424	1.8613	2.1257	2.6689
40	1.5676	1.8781	2.4692	1.8781	2.1477	2.6980
50	1.5770	1.8906	2.4883	1.8906	2.1635	2.7185
60	1.5844	1.9002	2.5026	1.9002	2.1755	2.7337
70	1.5906	1.9078	2.5137	1.9078	2.1849	2.7454
80	1.5958	1.9141	2.5225	1.9141	2.1925	2.7546
90	1.6002	1.9193	2.5297	1.9193	2.1986	2.7619
100	1.6040	1.9236	2.5355	1.9236	2.2037	2.7680
150	1.6171	1.9378	2.5536	1.9378	2.2198	2.7862
200	1.6249	1.9454	2.5623	1.9454	2.2279	2.7947
∞	1.6449	1.9600	2.5758	1.9600	2.2414	2.8070

*Fitted critical values are defined by $S_{N,\alpha}^+ = a_\alpha - b_\alpha \exp(-c_\alpha N^{1/2})$ and

$S_{N,\alpha} = A_\alpha - B_\alpha \exp(-C_\alpha N^{1/2})$, where the values of these coefficients are

α	a_α	b_α	c_α	A_α	B_α	C_α
0.10	1.6449	.23087	.12231	1.9600	.33036	.15599
0.05	1.9600	.33036	.15599	2.2414	.44989	.17537
0.01	2.5758	.56852	.18717	2.8070	.63821	.19757

FIGURE 1

Estimated critical values for the one-sided sequential KP procedure
at nominal levels .10, .05, and .01.

FIGURE 2

Estimated critical values for the two-sided sequential KP procedure
at nominal levels .10, .05, and .01.

The empirical critical values for the two-sided test at nominal level .10 were identical in all the cases that were simulated to the empirical critical values for the one-sided test at nominal level .05 (recall that the same set of simulated samples was used in the one-sided and two-sided cases). This is a reflection of the fact that, for these values, the second term on the right on (3.2) is negligible. Therefore, the fitted curve at the two-sided .10 level is identical to that at the one-sided .05 level (see Table V).

A simulation experiment was carried out to assess the accuracy of these fitted critical values for the values of N employed in the Monte Carlo comparison study to follow in Chapter 4. The results are summarized in Table VI. Each entry in the table is based on 1500 simulated samples (different from those which were used to construct Tables III and IV); entries for a fixed N but different values of p and α were obtained from the same set of simulated samples. Although considerable sample variability is evident in Table VI, it is nonetheless clear that the fitted critical values provide a substantial improvement over the asymptotic critical values.

Section 3.3 THE DAVIS, MGB AND MODIFIED MGB PROCEDURES, WITH EQUAL SAMPLE SIZES

For a fixed value of N , the set of simulated samples used to construct Tables III and IV was also used to determine empirical critical values for the other three sequential procedures. For the Davis procedure, curves of exactly the same form as described in Section 3.2 were fitted; a separate curve for the case $p = .25$ was used because of slower convergence of the empirical critical values to the asymptotic critical values. Davis indicated that the asymptotic critical values are useful for $Np > 60$, but curves were fit for the Davis procedure so that the Monte Carlo comparison study would

Table VI. Simulation Estimates of Actual Levels Corresponding to Fitted Critical Values for KP Procedures*.

m=n	P				
	.25	.50	.75	.90	1.0
1-sided case					
$\alpha = .10$					
15	.124	.105	.109	.107	.097
30	.080	.096	.097	.100	.097
50	.105	.114	.111	.113	.111
$\alpha = .05$					
15	.051	.052	.055	.049	.045
30	.053	.049	.050	.053	.052
50	.056	.055	.058	.054	.052
$\alpha = .01$					
15	.017	.015	.009	.008	.006
30	.009	.014	.013	.015	.011
50	.012	.010	.009	.014	.013
2-sided case					
$\alpha = .10$					
15	.108	.119	.111	.109	.099
30	.097	.100	.097	.101	.097
50	.105	.103	.097	.096	.092
$\alpha = .05$					
15	.053	.059	.056	.049	.043
30	.040	.053	.053	.053	.044
50	.053	.051	.057	.054	.049
$\alpha = .01$					
15	.009	.018	.013	.008	.007
30	.011	.011	.012	.011	.010
50	.008	.008	.011	.009	.007

*Each entry is based on 1500 simulated samples.

be as fair as possible. The situation was somewhat less tractable for the MGB and modified MGB procedures since asymptotic critical values and other appropriate asymptotic theory are unavailable for these procedures. The empirical critical values for the MGB and modified MGB procedures appeared to depend on N_p rather than N . Many different curves were tested, and eventually curves similar in form to those used by MGB were used to fit the empirical critical values as functions of N_p . The fitted curves are valid only for the range of N_p studied. The curves presented by MGB, used as functions of N_p , fitted the empirical critical values reasonably well, but MGB did not present curves for their one-sided test. The simulated samples which had been used to test the accuracy of the fitted critical values for the KP procedure were also used to test the accuracy of the fitted values for the Davis, MGB and modified MGB procedures. In all cases, the fitted critical values appeared adequate and were therefore used in the Monte Carlo comparison study.

Section 3.4 THE SEQUENTIAL PROCEDURES, WITH UNEQUAL SAMPLE SIZES

To this point we have restricted ourselves to the case of balanced experiments ($m=n$). Although the asymptotic critical values for both the KP and Davis procedures are also valid in the unbalanced case ($m \neq n$), larger sample sizes might be required before the asymptotic critical values provide useful approximations. Hence, it is of interest to examine the accuracy of the fitted critical values in the unbalanced case.

In the unbalanced case, with $\lambda = m/n$; the distributions of the modified Savage and Wilcoxon statistics are no longer symmetric about zero; calculations (Hájek and Šidák, 1967, p. 81-82) show that

$$E(SS_r^3) = (1 - 2\lambda) [N\lambda(1-\lambda)]^{1/2} g_s\left(\frac{r}{N}\right) + O(N^{-3/2}) \quad (3.5)$$

$$E(SW_r^3) = (1 - 2\lambda) [N\lambda(1-\lambda)]^{1/2} g_w\left(\frac{r}{N}\right) + O(N^{-3/2}) \quad (3.6)$$

where $g_s(p) = p^{-3/2} [2p + 3(1-p)\ln(1-p)]$, $g_w(p) = -3^{3/2} [1 - (1-p)^3]^{-3/2} p(1-p)^3$. $g_s(p)$ evaluated at $p = .25, .50, .75, .90, 1.0$ is respectively $-1.178, -.112, .709, 1.299, 2.000$; similarly, $g_w(p)$ evaluated at $p = .25, .50, .75, .90, 1.0$ is respectively $-1.247, -.397, -.062, -.005, 0.0$. Upon substituting these values into (3.5) and (3.6), one might be able to predict whether the critical values for the one-sided tests should be increased or decreased for different values of λ .

A simulation experiment was carried out to examine the accuracy of the fitted critical values in the unbalanced cases $\lambda = .25, .33, .40, .60, .67, .75$ for all the sequential procedures under consideration. For the one-sided tests, the results for the case $N = 100, \alpha = .05$ are summarized in Table VII, each entry is based on 1500 simulated samples. The results are similar for other values of N and α . For the one-sided MGB and modified MGB tests, there are obvious trends in the simulated significance levels which are increasing in λ . For the KP tests, there is a slight decreasing trend for $p = .75, .90$ and 1.0 . It appears that $g_s(p)$ and $g_w(p)$ are indicators of possible trends in the one-sided KP and Davis tests. For all the one-sided tests, there are slight trends in p which are increasing or decreasing depending on whether $\lambda < 1/2$ or $\lambda > 1/2$. These trends might be due to the fact that the moments (3.5) and (3.6) are increasing in p when $\lambda < 1/2$ and decreasing when $\lambda > 1/2$.

For the two-sided tests, the results for the case $N = 100, \alpha = .05$ are summarized in Table VIII; the test statistics have the same distribution

Table VII. Simulation Estimates of Actual Levels Corresponding to One-Sided Fitted Critical Values in the Unbalanced Case, with $N=100$, $\alpha = .05$.^{*}

	λ					
	.25	.33	.40	.60	.67	.75
<u>KP</u>						
p = .25	.047	.043	.050	.055	.046	.061
p = .50	.057	.041	.055	.055	.051	.046
p = .75	.057	.049	.057	.059	.042	.035
p = .90	.065	.047	.063	.051	.048	.028
p = 1.0	.062	.051	.063	.047	.043	.021
<u>MGB</u>						
p = .25	.020	.024	.037	.071	.072	.117
p = .50	.028	.028	.043	.075	.083	.113
p = .75	.029	.031	.044	.078	.084	.111
p = .90	.029	.031	.045	.076	.082	.085
p = 1.0	.030	.031	.046	.059	.066	.082
<u>Mod MGB</u>						
p = .25	.023	.034	.040	.068	.071	.117
p = .50	.037	.034	.046	.073	.078	.107
p = .75	.043	.035	.048	.073	.079	.078
p = .90	.045	.035	.049	.057	.078	.074
p = 1.0	.047	.035	.049	.056	.062	.071
<u>Davis</u>						
p = .25	.046	.045	.050	.062	.053	.061
p = .50	.053	.037	.047	.051	.050	.050
p = .75	.051	.044	.056	.055	.051	.042
p = .90	.053	.043	.055	.057	.052	.041
p = 1.0	.053	.044	.055	.057	.052	.041

* Each entry is based on 1500 simulated samples.

Table VIII. Simulation Estimates of Actual Levels Corresponding to Two-Sided Fitted Critical Values in the Unbalanced Case, with $N=100$, $\alpha = .05$.*

	$ \lambda - .5 $			
	0	.10	.17	.25
<u>KP</u>				
p = .25	.053	.052	.043	.049
p = .50	.051	.053	.048	.043
p = .75	.057	.054	.054	.050
p = .90	.054	.057	.052	.049
p = 1.0	.049	.052	.047	.046
<u>MGB</u>				
p = .25	.048	.048	.054	.061
p = .50	.055	.057	.058	.069
p = .75	.058	.060	.055	.072
p = .90	.049	.057	.057	.072
p = 1.0	.049	.057	.055	.071
<u>Mod MGB</u>				
p = .25	.049	.048	.059	.059
p = .50	.057	.057	.059	.066
p = .75	.058	.062	.055	.071
p = .90	.049	.058	.059	.072
p = 1.0	.048	.057	.058	.071
<u>Davis</u>				
p = .25	.055	.055	.047	.052
p = .50	.048	.051	.042	.041
p = .75	.049	.052	.044	.043
p = .90	.049	.051	.044	.043
p = 1.0	.049	.051	.044	.043

* Each entry in the first column is based on 1500 simulated samples, each entry in the other columns is based on 3000 simulated samples.

for λ and $1-\lambda$, therefore two estimates of the actual significance levels were averaged to obtain the entries in the table. The observed levels are reasonably close to the nominal level for the two-sided KP and Davis tests, while they tend to be high for the two-sided MGB and modified MGB tests, though not as extreme as for the one-sided case.

From this simulation experiment, it appears that the effect of unequal sample sizes on the level of the test is reasonably small for the KP and Davis procedures; the fitted critical values appear to be useful for the KP procedure provided $.33 \leq \lambda \leq .67$, and for the Davis procedure provided $.25 \leq \lambda \leq .75$. The effect of unequal sample sizes on the level of the test seems quite significant for the MGB and modified MGB procedures; particularly in the one-sided case where the fitted critical values appear to be useful only for λ near .5.

Section 3.5 THE FIXED POINT CENSORING PROCEDURES, WITH EQUAL SAMPLE SIZES

Although the comparison among the different sequential procedures is of primary interest, we will also compare the sequential procedures with their fixed point censoring counterparts. The standardized statistics for the fixed point censoring procedures (with censoring point r) based on the Savage, Wilcoxon and Mantel-Haenszel statistics are respectively $S_r^* / \{\text{var}(S_r^* | H_0)\}^{1/2}$ (or M_r^*), $W_r^* / \{\text{var}(W_r^* | H_0)\}^{1/2}$ and M_r^* . Each one of these statistics is asymptotically distributed as a standard normal random variable. We examined the accuracy of the asymptotic critical values for these fixed point censoring procedures in both the one-sided and two-sided cases using the set of simulated samples used to construct Tables III and IV. The asymptotic critical values appeared to be adequate in the Savage and Wilcoxon cases provided $N_p > 30$, but in the Mantel-Haenszel case they led to tests which rejected the null hypothesis too often (this last result

agrees with MGB, 1978, p. 619). MGB provided finite sample corrections to the asymptotic critical values in the case $p = 1$. Assuming that corrections should be similar for the other values of p , we tested the accuracy of these fitted critical values using the simulated samples used to construct Table VI. The fitted critical values seem adequate provided Np is not too small. Therefore, in the Monte Carlo comparison study, asymptotic critical values were used for the fixed point censoring procedures based on the Savage and Wilcoxon statistics, and the critical values provided by MGB were used for the fixed point censoring procedure based on the Mantel-Haenszel statistic.

Chapter 4

COMPARISON OF PROCEDURES

In this chapter, we compare the procedures under consideration in terms of both power and expected proportion sampled. We examine the properties of the procedures for the case of Lehmann alternatives specified by $1 - F_2(t) = [1 - F_1(t)]^\gamma$ for $0 < \gamma < \infty$. The null hypothesis is now $\gamma = 1$; the one-sided and two-sided alternatives are respectively $\gamma > 1$ and $\gamma \neq 1$. A simulation experiment was carried out to investigate these properties in the cases $m = n = 15, 30, 50$. The results for the one-sided procedures with $\alpha = .05$ are summarized in Table IX; each entry in the table is based on 500 simulated samples (The D-vectors can be simulated quite easily in the case of Lehmann alternatives, for example, see MGB, 1977, p. 624.). Results for the other values of α were qualitatively quite similar. The analogous results for the two-sided procedures are presented in Table X.

It was anticipated that, for larger sample sizes, the fixed point censoring procedures based on the Savage and Mantel-Haenszel statistics would have essentially identical power functions, since $\sum_{i=1}^r v_i$ is a consistent estimator of $\text{var}(S_r^* | H_0)$. It seems from the tables that this does hold, even for quite small sample sizes. The fixed point censoring procedure based on the Wilcoxon statistic appears to be essentially equivalent in power to the other two fixed point censoring procedures for the smaller values of p and less powerful than the other two procedures for the larger values of p .

Comparison of the sequential procedures is less straightforward than comparison of their fixed point censoring counterparts because both power and expected proportion sampled must be considered. The tables indicate that the

Table IX. Simulation Results for Operating Characteristics of the One-sided Tests Under Lehmann Alternatives, with $m = n = N/2$, $\alpha = .05^*$.

γ	<u>Fixed Point</u> <u>Censoring Power</u>			<u>Progressive Censoring</u> <u>Power</u>							
	S**	MH	W	KP	MGB	MMGB	D	Exp.	Prop.	Sampled	
<u>N = 30, p = .90</u>											
1.00	.07	.06	.05	.05	.06	.06	.05	.89	.87	.87	.88
1.25	.15	.15	.14	.14	.12	.12	.13	.87	.84	.84	.84
1.50	.27	.26	.26	.28	.25	.24	.26	.82	.77	.78	.77
2.00	.55	.54	.48	.52	.45	.42	.46	.74	.68	.69	.68
2.50	.74	.73	.68	.73	.60	.59	.62	.67	.60	.60	.60
3.00	.88	.87	.79	.84	.74	.72	.72	.61	.53	.53	.54
3.50	.91	.90	.84	.89	.83	.81	.81	.56	.46	.47	.48
<u>N = 30, p = 1.0</u>											
1.00	.06	.06	.05	.05	.06	.06	.05	.95	.93	.93	.94
1.25	.16	.16	.14	.13	.12	.12	.13	.93	.90	.90	.90
1.50	.28	.27	.26	.26	.24	.24	.26	.87	.83	.83	.82
2.00	.55	.55	.49	.51	.44	.42	.46	.79	.73	.73	.72
2.50	.75	.75	.68	.71	.60	.58	.62	.71	.63	.63	.63
3.00	.88	.88	.79	.83	.74	.72	.72	.63	.55	.55	.56
3.50	.91	.91	.84	.88	.82	.81	.81	.57	.48	.48	.49
<u>N = 60, p = .50</u>											
1.00	.03	.03	.03	.04	.03	.03	.03	.50	.49	.49	.49
1.20	.12	.12	.13	.14	.10	.10	.12	.48	.47	.47	.48
1.35	.20	.19	.20	.20	.16	.16	.18	.48	.46	.46	.47
1.50	.30	.28	.29	.31	.23	.23	.27	.46	.45	.45	.45
1.75	.38	.37	.37	.38	.30	.30	.34	.45	.43	.43	.44
2.00	.59	.58	.57	.60	.49	.49	.55	.41	.39	.39	.40
2.25	.70	.69	.70	.71	.61	.60	.65	.39	.36	.36	.38

*Each Entry is based on 500 simulated samples.

**S = Savage; MH = Mantel-Haenszel; W = Wilcoxon; KP = Koziol and Petkau; MGB = Muenz, Green and Byar; MMGB = modified MGB, D = Davis.

Table IX. (continued)

γ	S	MH	W	KP	MGB	MMGB	D	KP	MGB	MMGB	D
<u>N = 60, p = .75</u>											
1.00	.04	.04	.03	.04	.03	.03	.04	.74	.74	.74	.74
1.20	.16	.15	.14	.14	.12	.12	.12	.72	.70	.70	.71
1.35	.23	.22	.21	.23	.18	.18	.20	.70	.67	.67	.68
1.50	.38	.37	.37	.36	.30	.29	.32	.67	.64	.64	.64
1.75	.54	.53	.49	.53	.42	.41	.44	.63	.59	.60	.61
2.00	.74	.74	.70	.72	.63	.62	.66	.56	.51	.51	.52
2.25	.83	.82	.78	.81	.73	.73	.76	.52	.46	.46	.47
<u>N = 60, p = .90</u>											
1.00	.05	.05	.03	.04	.04	.03	.04	.89	.88	.88	.88
1.20	.18	.18	.15	.18	.13	.12	.13	.86	.83	.83	.84
1.35	.27	.27	.23	.27	.19	.18	.20	.83	.80	.80	.81
1.50	.43	.43	.38	.41	.32	.31	.33	.79	.75	.75	.75
1.75	.62	.62	.53	.62	.46	.46	.46	.72	.68	.68	.69
2.00	.79	.79	.71	.78	.68	.67	.68	.64	.57	.57	.57
2.25	.89	.89	.82	.87	.79	.78	.77	.57	.50	.50	.51
<u>N = 60, p = 1.0</u>											
1.00	.05	.05	.03	.04	.03	.03	.04	.97	.96	.96	.96
1.20	.19	.19	.15	.17	.13	.12	.13	.93	.91	.91	.92
1.35	.29	.29	.22	.25	.18	.18	.20	.90	.87	.87	.87
1.50	.45	.46	.38	.42	.32	.31	.33	.85	.80	.81	.80
1.75	.63	.64	.54	.60	.47	.45	.46	.77	.73	.73	.74
2.00	.81	.82	.72	.78	.67	.66	.68	.67	.59	.60	.60
2.25	.90	.90	.82	.87	.79	.78	.77	.60	.52	.52	.53
<u>N = 100, p = .25</u>											
1.00	.06	.06	.06	.06	.05	.05	.07	.25	.24	.24	.25
1.15	.08	.08	.09	.09	.08	.08	.10	.24	.24	.24	.24
1.25	.15	.14	.15	.14	.10	.10	.13	.24	.24	.24	.24
1.50	.26	.26	.26	.26	.22	.22	.27	.23	.23	.23	.23
1.75	.37	.36	.38	.37	.28	.28	.37	.23	.22	.22	.23
2.00	.51	.50	.51	.50	.43	.43	.52	.22	.20	.20	.21

Table IX. (continued)

γ	S	MH	W	KP	MGB	MMGB	D	KP	MGB	MMGB	D
<u>N = 100, p = .50</u>											
1.00	.06	.06.	.07	.05	.06	.06	.05	.49	.48	.48	.49
1.15	.14	.14	.13	.13	.11	.12	.12	.49	.47	.47	.48
1.25	.20	.20	.20	.19	.16	.16	.18	.47	.46	.46	.47
1.50	.39	.38	.38	.38	.32	.32	.34	.45	.41	.41	.43
1.75	.62	.62	.62	.59	.45	.46	.53	.41	.39	.38	.40
2.00	.81	.81	.80	.79	.66	.66	.75	.37	.33	.33	.35
<u>N = 100, p = .75</u>											
1.00	.06	.06	.06	.06	.05	.05	.05	.74	.72	.72	.73
1.15	.14	.14	.15	.15	.13	.13	.14	.72	.69	.69	.71
1.25	.26	.26	.24	.24	.20	.20	.21	.70	.67	.67	.68
1.50	.54	.54	.48	.52	.39	.39	.43	.63	.59	.59	.61
1.75	.76	.76	.73	.75	.60	.60	.65	.56	.52	.52	.52
2.00	.89	.89	.87	.89	.83	.82	.84	.48	.41	.41	.42
<u>N = 100, p = .90</u>											
1.00	.06	.06	.06	.06	.05	.06	.05	.89	.87	.86	.87
1.15	.15	.15	.14	.15	.14	.13	.14	.86	.83	.83	.84
1.25	.30	.30	.24	.27	.20	.19	.21	.83	.80	.80	.80
1.50	.58	.59	.53	.59	.45	.44	.46	.73	.68	.68	.69
1.75	.84	.84	.77	.79	.68	.67	.69	.64	.58	.58	.58
2.00	.94	.04	.90	.92	.86	.85	.85	.53	.44	.44	.45
<u>N = 100, p = 1.0</u>											
1.00	.05	.05	.07	.06	.05	.05	.05	.98	.95	.95	.96
1.15	.16	.16	.15	.15	.13	.13	.14	.94	.91	.91	.91
1.25	.28	.29	.25	.26	.19	.19	.21	.91	.87	.87	.87
1.50	.63	.64	.53	.58	.45	.45	.46	.79	.73	.73	.74
1.75	.86	.87	.77	.81	.69	.67	.69	.67	.61	.62	.61
2.00	.95	.95	.90	.93	.87	.86	.85	.56	.45	.45	.46

Table X. Simulation Results for Operating Characteristics of the Two-sided Tests Under Lehmann Alternatives, with $m = n = N/2$, $\alpha = .05^*$.

γ	<u>Fixed Point</u> <u>Censoring Power</u>			<u>Progressive Censoring</u>							
	S**	MH	W	<u>Power</u>				<u>Exp. Prop. Sampled</u>			
	KP	MGB	MMGB	D	KP	MGB	MMGB	D			
<u>N = 30, p = .90</u>											
1.00	.06	.06	.05	.06	.06	.05	.06	.89	.87	.87	.88
1.25	.09	.09	.08	.09	.07	.06	.07	.88	.87	.87	.87
1.50	.18	.18	.21	.19	.17	.15	.20	.85	.82	.82	.81
2.00	.42	.42	.38	.41	.33	.31	.34	.79	.75	.76	.75
2.50	.59	.59	.53	.56	.47	.45	.49	.73	.69	.69	.69
3.00	.75	.75	.66	.73	.63	.60	.62	.67	.62	.62	.62
3.50	.84	.83	.77	.81	.74	.72	.72	.62	.55	.55	.55
<u>N = 30, p = 1.0</u>											
1.00	.05	.05	.05	.06	.06	.05	.06	.95	.94	.94	.94
1.25	.08	.09	.08	.07	.07	.06	.07	.95	.93	.93	.94
1.50	.18	.18	.21	.17	.17	.15	.20	.91	.87	.88	.87
2.00	.41	.42	.38	.39	.33	.29	.34	.84	.80	.81	.80
2.50	.58	.59	.53	.53	.47	.43	.49	.78	.72	.73	.72
3.00	.75	.76	.66	.70	.62	.60	.62	.71	.64	.65	.65
3.50	.83	.84	.77	.80	.73	.70	.72	.64	.57	.58	.57
<u>N = 60, p = .50</u>											
1.00	.04	.03	.03	.03	.03	.03	.03	.50	.49	.49	.50
1.20	.08	.07	.07	.08	.06	.06	.07	.49	.49	.49	.49
1.35	.13	.12	.14	.14	.11	.10	.12	.49	.48	.48	.48
1.50	.18	.18	.18	.18	.16	.16	.17	.48	.46	.46	.47
1.75	.28	.26	.26	.28	.20	.20	.24	.47	.46	.46	.47
2.00	.48	.46	.46	.47	.36	.36	.41	.44	.42	.42	.43
2.25	.59	.59	.58	.60	.49	.49	.55	.42	.40	.40	.41

*Each Entry is based on 500 simulated samples.

**S = Savage; MH = Mantel-Haenszel; W = Wilcoxon; KP = Koziol and Petkau;
MGB = Muenz, Green and Byar; MMGB = modified MGB, D = Davis.

Table X. (continued)

γ	S	MH	W	KP	MGB	MMGB	D	KP	MGB	MMGB	D
<u>N = 60, p = .75</u>											
1.00	.05	.05	.04	.05	.04	.04	.04	.74	.74	.74	.74
1.20	.09	.09	.09	.10	.07	.07	.08	.73	.72	.72	.72
1.35	.15	.15	.15	.15	.12	.12	.13	.72	.70	.70	.71
1.50	.28	.27	.24	.28	.22	.22	.23	.70	.67	.67	.68
1.75	.43	.42	.37	.42	.32	.31	.34	.67	.65	.65	.65
2.00	.62	.61	.59	.61	.51	.50	.53	.61	.57	.57	.59
2.25	.73	.73	.70	.74	.64	.63	.66	.57	.52	.52	.53
<u>N = 60, p = .90</u>											
1.00	.05	.05	.04	.05	.04	.04	.04	.89	.88	.88	.88
1.20	.11	.11	.10	.11	.09	.08	.08	.88	.86	.86	.86
1.35	.17	.17	.15	.16	.14	.13	.14	.86	.84	.84	.84
1.50	.30	.30	.26	.30	.24	.24	.23	.82	.79	.79	.80
1.75	.48	.49	.38	.47	.35	.33	.35	.78	.75	.75	.75
2.00	.70	.70	.62	.67	.57	.53	.55	.70	.65	.65	.66
2.25	.80	.80	.72	.79	.68	.66	.68	.64	.58	.58	.58
<u>N = 60, p = 1.0</u>											
1.00	.05	.06	.04	.05	.04	.04	.04	.97	.96	.96	.96
1.20	.12	.12	.09	.10	.08	.08	.08	.95	.94	.94	.94
1.35	.18	.18	.16	.15	.13	.13	.14	.94	.91	.91	.91
1.50	.33	.34	.26	.29	.24	.23	.23	.90	.85	.86	.87
1.75	.49	.50	.39	.44	.34	.32	.35	.84	.81	.81	.81
2.00	.70	.72	.62	.66	.56	.53	.55	.74	.69	.69	.70
2.25	.82	.83	.72	.79	.68	.65	.68	.68	.61	.61	.61
<u>N = 100, p = .25</u>											
1.00	.06	.06	.06	.05	.05	.05	.06	.25	.24	.24	.25
1.15	.07	.07	.07	.06	.05	.05	.06	.25	.24	.24	.25
1.25	.11	.11	.09	.08	.07	.07	.09	.25	.24	.24	.24
1.50	.20	.20	.18	.18	.14	.14	.20	.24	.23	.23	.24
1.75	.30	.29	.27	.24	.18	.19	.25	.24	.23	.23	.24
2.00	.42	.42	.40	.35	.29	.29	.37	.23	.22	.22	.23

Table X. (continued)

γ	S	MH	W	KP	MGB	MMGB	D	KP	MGB	MMGB	D
<u>N = 100, p = .50</u>											
1.00	.03	.03	.04	.04	.05	.05	.04	.50	.49	.49	.49
1.15	.08	.08	.08	.07	.06	.07	.07	.49	.48	.48	.49
1.25	.12	.12	.12	.13	.11	.11	.11	.48	.47	.47	.48
1.50	.28	.27	.26	.27	.22	.22	.24	.47	.44	.44	.46
1.75	.46	.46	.45	.44	.34	.34	.40	.44	.42	.42	.44
2.00	.70	.70	.68	.69	.55	.55	.62	.40	.38	.37	.39
<u>N = 100, p = .75</u>											
1.00	.05	.05	.05	.05	.05	.05	.04	.74	.73	.73	.74
1.15	.10	.10	.09	.10	.08	.08	.08	.73	.72	.72	.73
1.25	.18	.18	.17	.17	.13	.13	.14	.72	.70	.70	.71
1.50	.41	.41	.38	.39	.30	.30	.32	.67	.63	.63	.65
1.75	.66	.66	.59	.64	.50	.50	.54	.61	.57	.57	.59
2.00	.85	.84	.81	.83	.74	.73	.77	.53	.47	.47	.49
<u>N = 100, p = .90</u>											
1.00	.04	.05	.05	.04	.05	.05	.04	.89	.87	.87	.88
1.15	.09	.10	.09	.09	.09	.09	.08	.88	.86	.86	.87
1.25	.19	.19	.18	.19	.15	.15	.15	.86	.83	.83	.84
1.50	.48	.48	.41	.47	.35	.34	.34	.79	.74	.74	.76
1.75	.72	.73	.65	.71	.57	.56	.55	.70	.65	.66	.66
2.00	.89	.90	.83	.89	.80	.80	.80	.60	.52	.52	.53
<u>N = 100, p = 1.0</u>											
1.00	.04	.04	.04	.04	.05	.05	.04	.98	.96	.96	.97
1.15	.09	.10	.09	.09	.09	.09	.09	.97	.94	.94	.95
1.25	.20	.20	.18	.19	.15	.15	.15	.94	.91	.91	.91
1.50	.49	.51	.41	.46	.35	.34	.34	.85	.80	.81	.81
1.75	.75	.77	.65	.72	.58	.56	.56	.74	.70	.70	.70
2.00	.91	.91	.84	.89	.80	.79	.80	.63	.54	.54	.55

MGB and modified MGB procedures are essentially equivalent and that the Davis procedure tends to dominate both of them. The three procedures are equivalent in power near $p=1$, but the Davis procedure is more powerful than the MGB and modified MGB procedures when p is removed from 1. Although the MGB procedure appears to have a slight advantage in terms of expected proportion sampled, the differences with the modified MGB and Davis procedures are negligible. Thus, on the basis of the cases considered, one might prefer the Davis procedure to the MGB and modified MGB procedures. The KP procedure is more powerful than the Davis procedure; the differences in power being larger in the cases where p is close to 1. On the other hand, the Davis procedure has the advantage in terms of expected proportion sampled, and this advantage is most pronounced in those cases where the KP procedure has the greatest advantage in power. In choosing between these two sequential procedures in any practical situation, the relative importance of power and expected proportion sampled should be taken into consideration.

It is also of interest to compare the sequential procedures to their fixed point censoring counterparts; the sequential procedures are, of course, less powerful. The tables indicate that the MGB procedure suffers the most loss in power relative to its fixed point censoring counterpart. One might not be willing to tolerate such losses in power for gains in expected proportion sampled. The losses in power for the Davis procedure are more reasonable, while for the KP procedure the losses in power are quite small. Of course, all the sequential procedures offer substantial gains in the expected proportion sampled, particularly for the case in which there are large differences between the two treatments under consideration.

For the special case of contiguous exponential scale shift alternatives, Koziol and Petkau (1978; see also Lesser, 1979) were able to determine the

asymptotic properties of both the fixed point and progressive censoring procedures based on the Savage statistic. While the asymptotic results do not provide accurate approximations for the sample sizes under consideration, the conclusions to which one is led on the basis of the asymptotics are in fact verified by the Monte Carlo results; that is, for the procedures based on the Savage statistic, while the power of the sequential procedure is only slightly smaller than that of the fixed point censoring procedure, substantial reductions in the expected proportion sampled can result from the use of the sequential procedure.

In summary, it appears that, at least in the case of Lehmann alternatives, the choice is narrowed to the fixed point censoring procedure based on the Savage statistic (or the equivalent procedure based on the Mantel-Haenszel statistic) and the sequential procedures based on the Savage and Wilcoxon statistics. If the situation is such that maximal power is essential, one would use the fixed point censoring procedure. If one is willing to entertain minimal losses in power in exchange for substantial reduction in expected proportion sampled, one would use the KP procedure. Somewhat marginal further gains in expected proportion sampled can be achieved at the expense of losses in power by the use of the Davis procedure.

Chapter 5

EARLY DECISION RULES

The results in Chapter 4 show that use of the sequential procedures can result in substantial time savings. Further time savings can be achieved through the use of early decision rules; this will be investigated in this chapter. By an early decision rule we mean a rule for terminating the experiment which has the property that if the experiment had been continued (to the limits imposed by the design), the decision made would have been no different. While such rules can be expected to be most effective when used in conjunction with fixed point censoring procedures, they can also be expected to improve the behaviour of the sequential procedures in those instances where one would usually be led to accept the null hypothesis.

Consider the modified form of a linear rank statistic based on the non-decreasing scores $\{a_N(i)\}$ and defined by

$$T_k^* = \sum_{i=1}^k d_i a_N(i) + (m - m_k) \bar{a}_N(k) \quad (5.1)$$

where $m_k = \sum_{i=1}^k d_i$ and $\bar{a}_N(k) = \sum_{j=k+1}^N a_N(j) / (N-k)$. Also let $n_k = k - m_k$. Note that $T_{N-1}^* = T_N^* = T_N$, the usual form of the linear rank statistic. Now, for $k < N-1$,

$$T_{k+1}^* = T_k^* - [d_{k+1} - (m - m_k) / (N-k)] [\bar{a}_N(k+1) - a_N(k+1)].$$

Since $\bar{a}_N(k+1) - a_N(k+1) \geq 0$, upper and lower bounds for T_{k+1}^* in terms of T_k^* are obtained by setting d_{k+1} equal to 0 and 1 respectively in this expression.

Noting further that the value of T_k^* no longer changes (with k) after all the order statistics corresponding to either group have been observed, it is easy to determine upper and lower bounds on the possible values of T_r^* ($r > k$) in terms of m_k and T_k^* . In particular

$$L_r(k) \leq T_r^* \leq U_r(k)$$

where

$$\begin{aligned} U_r(k) &= T_k^* + (m - m_k) [\bar{a}_N(r) - \bar{a}_N(k)] \quad \text{if } n - n_k \geq r - k \\ &= T_k^* + (m - m_k) [\bar{a}_N(n + m_k) - \bar{a}_N(k)] \quad \text{if } n - n_k < r - k \end{aligned}$$

and

$$\begin{aligned} L_r(k) &= T_k^* - (n - n_k) [\bar{a}_N(r) - \bar{a}_N(k)] \quad \text{if } m - m_k \geq r - k \\ &= T_k^* - (n - n_k) [\bar{a}_N(m + n_k) - \bar{a}_N(k)] \quad \text{if } m - m_k < r - k. \end{aligned}$$

These bounds can now be used to define early decision rules to be used in conjunction with both the fixed point and progressive censoring procedures based on $\{a_N(i)\}$. In particular, for the one-sided test based on T_r^* , the early decision rule is as follows. For each k , $1 \leq k \leq r$, compute $L_r(k)$ and $U_r(k)$. Stop and accept H_0 at the first k for which $U_r(k)$ is less than or equal to the critical value for the test based on T_r^* ; stop and reject H_0 at the first k for which $L_r(k)$ is greater than this same critical value.

Similarly, for the one-sided test based on $\max_{1 \leq k \leq r} T_k^*$, the early decision rule

is as follows. For each k , $1 \leq k \leq r$, compute $U_r(k)$. Stop and accept H_0 at the first k for which $U_r(k)$ is less than or equal to the critical value for the test based on $\max_{1 \leq k \leq r} T_k^*$; stop and reject H_0 at the first k for which T_k^*

exceeds this same critical value. The early decision rules are defined similarly in the two-sided case. The early decision rule based on $|T_r^*|$ is to accept H_0 at the first k for which $\max\{|L_r(k)|, |U_r(k)|\}$ is less than or equal to the critical value for the test based on $|T_r^*|$, and to reject H_0 at the first k for which either $L_r(k)$ or $-U_r(k)$ is greater than this critical value. The early decision rule based on $\max_{1 \leq k \leq r} |T_k^*|$ is to accept H_0 at the first k for which $\max\{|L_r(k)|, |U_r(k)|\}$ is less than or equal to the critical value for the test based on $\max_{1 \leq k \leq r} |T_k^*|$, and to reject H_0 at the first k for which $|T_k^*|$ exceeds this same critical value. Since both the Savage and Wilcoxon statistics are of form (5.1), the above serves to define early decision rules for the fixed point and progressive censoring procedures based on these statistics. On the other hand, the Mantel-Haenszel statistics M_k^* are not of form (5.1). It appears, however, that (for $1 \leq r < N$) $M_{r+1}^* \geq M_r^*$ if $d_{r+1} = 0$ and $M_{r+1}^* \leq M_r^*$ if $d_{r+1} = 1$. If this were the case, early decision rules of the general form above could be defined for the fixed point and progressive censoring procedures based on the Mantel-Haenszel statistics. The appropriate upper bound, $U_r(k)$, would be the value M_r^* would achieve based on the D-sequence which has been observed up to and including the k th order statistic and extended by adding first all the remaining 0's followed by the remaining 1's. The appropriate lower bound, $L_r(k)$, would be determined similarly with the 1's added before the 0's. MGB, in fact, use these bounds in the case $r = N$ for their early decision rules. For the modified MGB statistics, M_k' , it appears that $M_{r+1}' \geq M_r'$ if $d_{r+1} = 0$ and $M_{r+1}' \leq M_r'$ if $d_{r+1} = 1$, provided not all of the order statistics have been observed in any one of the two groups. If so, bounds $U_r(k)$ and $L_r(k)$ can be found, and early decision rules exist for the modified MGB procedure.

In order to examine the gains provided by these early decision rules, the cases (with $p = 1$) displayed in Tables IX and X were redone (using the same simulations) for the procedures based on the Savage, Wilcoxon and Mantel-Haenszel statistics. The resulting values of the expected proportions sampled are summarized in Table XI and XII. It is clear that while substantial gains are possible for the fixed point censoring procedures, most of the further gains for the sequential procedures are made in the neighbourhood of the null hypothesis. Of course, any savings in time which can be achieved at no expense in power are noteworthy. If one were planning to use one of these fixed point or progressive censoring procedures, the early decision rules should be employed.

Table XI. Expected Proportions Sampled Using Early Decision Rules in the One-Sided Case, with $\alpha = .05$, $p = 1^*$.

<u>Fixed Point Censoring</u>				<u>Progressive Censoring</u>		
γ	S**	MH	W	KP	MGB	D
<u>N = 30</u>						
1.00	.59	.61	.45	.55	.47	.40
1.25	.67	.68	.52	.61	.52	.45
1.50	.71	.72	.55	.62	.51	.45
2.00	.77	.77	.61	.64	.52	.47
2.50	.77	.78	.61	.62	.49	.47
3.00	.76	.76	.60	.58	.47	.45
3.50	.72	.72	.58	.54	.42	.42
<u>N = 60</u>						
1.00	.70	.71	.53	.66	.59	.48
1.20	.79	.80	.61	.71	.63	.53
1.35	.80	.81	.63	.71	.61	.53
1.50	.84	.84	.66	.72	.61	.53
1.75	.85	.85	.68	.69	.60	.54
2.00	.83	.83	.66	.63	.52	.48
2.25	.81	.80	.64	.58	.47	.45
<u>N = 100</u>						
1.00	.78	.78	.59	.73	.65	.53
1.15	.82	.83	.65	.75	.67	.56
1.25	.86	.86	.68	.77	.67	.57
1.50	.88	.89	.72	.72	.63	.56
1.75	.86	.86	.70	.65	.56	.51
2.00	.82	.82	.65	.55	.43	.42

* Each entry is based on 500 simulated samples.

** S = Savage, MH = Mantel-Haenszel, W = Wilcoxon,
 KP = Koziol and Petkau, MGB = Muenz, Green and Byar,
 D = Davis.

Table XII. Expected Proportions Sampled Using Early Decision Rules in the Two-Sided Case, with $\alpha = .05$, $p = 1^*$.

<u>Fixed Point Censoring</u>				<u>Progressive Censoring</u>		
γ	S**	MH	W	KP	MGB	D
<u>N = 30</u>						
1.00	.64	.66	.52	.60	.54	.47
1.25	.65	.67	.53	.61	.54	.47
1.50	.69	.71	.56	.63	.54	.48
2.00	.73	.74	.59	.64	.55	.49
2.50	.76	.76	.62	.64	.54	.50
3.00	.76	.76	.61	.63	.52	.48
3.50	.73	.73	.59	.59	.48	.46
<u>N = 60</u>						
1.00	.75	.76	.58	.71	.64	.53
1.20	.77	.78	.60	.72	.64	.54
1.35	.79	.80	.62	.73	.64	.54
1.50	.81	.82	.65	.73	.63	.55
1.75	.84	.84	.67	.72	.63	.56
2.00	.84	.84	.67	.68	.58	.53
2.25	.82	.82	.65	.64	.53	.50
<u>N = 100</u>						
1.00	.81	.83	.64	.77	.70	.59
1.15	.82	.83	.66	.77	.70	.60
1.25	.84	.85	.67	.78	.69	.59
1.50	.88	.88	.71	.76	.67	.59
1.75	.88	.87	.72	.70	.62	.56
2.00	.84	.83	.68	.61	.51	.49

* Each entry is based on 500 simulated samples.

** S = Savage, MH = Mantel-Haenszel, W = Wilcoxon,
 KP = Koziol and Petkau, MGB = Muenz, Green and Byar,
 D = Davis.

Chapter 6

EXAMPLES

Section 6.1 THE FIRST EXAMPLE

MGB consider a hypothetical trial to investigate the effect on survival of a new adjuvant therapy for lung cancer, the therapy to be administered at the time of initial treatment. Thirty-two patients are initially enrolled in the study, 16 in the control group (standard therapy) and 16 in the test group (standard plus adjuvant). The D-vector corresponding to the ranked lengths of survival is (000001000100011011111111010111001), where 1 denotes a patient from the control group, and 0 denotes a patient from the test group. We wish to test the equality of the underlying survival distributions against the two-sided alternative, using the various procedures discussed previously.

The statistics $S_k^*/\{\text{var}(S_N|H_0)\}^{1/2}$, $W_k^*/\{\text{var}(W_N|H_0)\}^{1/2}$, M_k^* and M_k' ($1 \leq k \leq 32$) are tabulated in Table XIII.

The values of the test statistics in the uncensored case are $|S_N|/\{\text{var}(S_N|H_0)\}^{1/2} = 1.502$, $|W_N|/\{\text{var}(W_N|H_0)\}^{1/2} = 2.337$ and $M_N = 1.514$; the critical values at the $\alpha = .05$ level are approximately 1.96, 1.96 and 2.06 respectively. From these values, one can see that the Wilcoxon statistic is more sensitive to departures from the null hypothesis evidenced in the data than either the Savage or Mantel-Haenszel statistics; only the test based on the Wilcoxon statistic will reject H_0 at the $\alpha = .05$ level. However, each of the sequential procedures (with $p = 1$) will reject H_0 at the $\alpha = .05$ level. The critical values for the KP, Davis, MGB and modified MGB test statistics are approximately 2.07, 2.12, 2.60 and 2.60 respectively. Thus, the KP procedure will reject H_0 at the 13th observation and the Davis,

Table XIII Statistics for the First Example

k	d _k	S' _k	W' _k	M _k [*]	M _k ⁱ
1	0	0.186	0.302	1.000	1.000
2	0	0.378	0.603	1.437	1.437
3	0	0.577	0.905	1.791	1.790
4	0	0.782	1.206	2.105	2.103
5	0	0.995	1.508	2.399	2.393
6	1	0.844	1.300	1.861	1.852
7	0	1.058	1.583	2.163	2.152
8	0	1.282	1.866	2.455	2.439
9	0	1.515	2.148	2.742	2.718
10	1	1.385	1.998	2.387	2.359
11	0	1.622	2.261	2.671	2.635
12	0	1.870	2.525	2.959	2.910
13	0	2.131	2.789	3.255	3.187
14	1	2.033	2.695	3.012	2.932
15	1	1.930	2.601	2.775	2.690
16	0	2.192	2.827	3.063	2.961
17	1	2.099	2.751	2.862	2.752
18	1	2.000	2.676	2.660	2.550
19	1	1.894	2.601	2.458	2.351
20	1	1.779	2.525	2.254	2.155
21	1	1.655	2.450	2.046	1.958
22	1	1.519	2.374	1.834	1.758
23	1	1.371	2.299	1.616	1.553
24	0	1.577	2.393	1.817	1.752
25	1	1.438	2.337	1.622	1.567
26	0	1.651	2.412	1.823	1.767
27	1	1.526	2.374	1.655	1.607
28	1	1.377	2.337	1.465	1.427
29	1	1.191	2.299	1.243	1.217
30	0	1.315	2.318	1.350	1.327
31	0	1.502	2.337	1.514	1.502
32	1	1.502	2.337	1.514	1.502

Note: $S'_k = S_k^* / \{\text{var}(S_N | H_0)\}^{1/2}$, $W'_k = W_k^* / \{\text{var}(W_N | H_0)\}^{1/2}$

MGB and modified MGB procedures will reject H_0 at the 9th observation. In this example, the decisions resulting from the KP and MGB procedures are different (at the $\alpha = .05$ level) from those resulting from their fixed point censoring counterparts. There is this risk involved in one wishes to make a decision with incomplete survival information.

It is not surprising that the Davis procedure rejects H_0 earlier than the KP procedure, given the relative performances of the Wilcoxon and Savage statistics with the overall survival data. That the MGB procedure rejects H_0 before the KP procedure is to be expected here, given the nature of the two procedures (see KP, 1978, p. 622). Consider the scheme of monitoring the course of the trial by plotting the points (k, S_k^*) sequentially and connecting by straight lines. Under H_0 , the resulting curve will vary about the x-axis. The MGB and KP procedures impose square root and horizontal boundaries respectively on this graph; H_0 would be rejected (and the trial would cease) were the curve to cross the appropriate boundary. In this example, the consistent pattern of early deaths from the test group will lead to earlier rejection of H_0 by the MGB procedure than the KP procedure, since the square root boundary is initially within the horizontal boundary.

Section 6.2 THE SECOND EXAMPLE

Martin (1946) measured the length of survival of mice infected with tubercle bacilli; his experimental results (Mosteller and Tukey, 1977, p. 506) are given as follows in Table XIV. The total sample size is 362; the size of group A is 226, the size of group B is 138.

Table XIV Mouse Survival Data

Group	Survival (in days after inoculation)								
	14+	15+	16+	17+	18+	19+	20+	21+	22+
A	7	6	12	23	23	43	37	23	17
B	1	0	0	3	2	7	12	14	16
Group	23+	24+	25+	26+	27+	28+	29+	30+	≥ 31
A	6	7	8	5	1	2	1	0	3
B	16	24	12	4	0	0	4	4	19

We wish to test the equality of the underlying survival distributions of the two groups against the two-sided alternative, using the various procedures discussed previously. Randomization was used to assign ranks to tied observations; this technique precludes any changes in the levels of significance of the various procedures attributable to the discreteness of the data (Hájek and Šidák, 1967, Section III.8.1).

The values of the test statistics in the uncensored case are $|S_N|/\{\text{var}(S_N|H_0)\}^{1/2} = 8.892$, $|W_N|/\{\text{var}(W_N|H_0)\}^{1/2} = 9.791$ and $|M_N| = 9.321$. Because these values are so highly significant ($P_{\text{obs}} < .001$), one can conclude that the two underlying survival distributions are not identical. The sequential procedures (with $p = 1$) would have terminated the study quite early (at the $\alpha = .05$ level); the KP procedures would have rejected H_0 at the 60th observation, the Davis procedure at the 37th observation, and the MGB and modified MGB procedures at the 19th observation. In terms of time, the procedures would reject H_0 between the 16th and 18th days (inclusive) of the study; the time saving afforded by the sequential procedures is substantial.

Remark. In these two examples, the use of the early decision rules described in Chapter 5 would not affect the sequential procedures (at the $\alpha = .05$ level). Use of the early decision rules (at the $\alpha = .05$ level) would result in early decisions for the uncensored tests. In the first example, the tests based on the Savage and Mantel-Haenszel statistics would accept H_0 at the 28th observation and the test based on the Wilcoxon statistic would reject H_0 at the 24th observation. In the second example, the tests based on the Savage, Wilcoxon and Mantel-Haenszel statistics would reject H_0 at the 219th, 151st and 216th observations respectively.

Chapter 7

DISCUSSION

The simulation studies show that of the procedures considered in this thesis, the "best" procedures (under Lehmann alternatives) are the fixed point censoring procedure based on the Savage statistic, the sequential KP procedure and the sequential Davis procedure. If power against Lehmann alternatives is of predominant concern, the fixed point censoring procedure based on the Savage statistic is to be preferred. If the expected proportion sampled is of major concern, the Davis procedure is to be preferred. A good compromise is the KP procedure, which loses little power with respect to the fixed point censoring procedure and is only marginally worse in terms of expected proportion sampled with respect to the Davis procedure. In a clinical trial setting, where the decision to be made will affect future patients, it would appear that power is of primary importance, and one would be willing to sacrifice small losses in power only for substantial gains in the expected proportion sampled. From this point of view, one might prefer the KP procedure. If one decided to use the KP procedure, the critical values presented in Table V should be used, as use of the asymptotic critical values for the KP statistics result in somewhat conservative tests (at the significance levels of usual interest), even for moderate-sized samples.

The other two sequential procedures under consideration, the MGB and modified MGB procedures, tend to be dominated by the Davis procedure, the Davis procedure is generally more powerful and is negligibly different in terms of expected proportion sampled (with the early decision rules, the Davis procedure appears to be superior in terms of expected proportion sampled, see Tables XI and XII). The MGB and modified MGB procedures also have the

disadvantages of having no asymptotic theory and having critical values which change considerably (at least for the one-sided test) as $\lambda = m/N$ varies from $1/2$).

For any of the suggested procedures above, the early decision rules discussed in Chapter 5 can be profitably incorporated. There are gains in expected proportion sampled, together with no loss of power. Some additional computational labour is required for its implementation, however. Also, for the sequential procedures, essentially no gains in power accrue as p increases from $.9$ to 1 ; exploitation of this observation would lead to time savings with essentially no attendant loss of power.

If an experiment were to be truncated at a particular prespecified point in time T , rather than at an order statistic, the results in this thesis still hold; "r out of N " censoring is asymptotically equivalent (under H_0) to truncation at T , when $F_1^{-1}(r/N) \rightarrow T$ as $N \rightarrow \infty$ (cf. Chatterjee and Sen, 1973). Further information on the analysis of survival data in clinical trials can be found in Armitage (1975) and Peto et al (1976, 1977).

BIBLIOGRAPHY

- Armitage, P. (1975). Sequential Medical Trials. 2nd Ed. John Wiley, New York.
- Bickel, P.J. and van Zwet, W.R. (1978). Asymptotic expansions for the power of distribution-free tests in the two-sample problem. Ann. Statist. 6, 937-1004.
- Chatterjee, S.K. and Sen, P.K. (1973). Nonparametric testing under progressive censorship. Cal. Statist. Bull. 22, 13-50.
- Davis, G.E. (1978). A two-sample Wilcoxon test for progressively censored data. Commun. Statist. Theor. Meth. A7(4), 389-98.
- Gastwirth, J.L. (1965). Asymptotically most powerful rank tests for the two-sample problem with censored data. Ann. Math. Statist. 36, 1243-7.
- Hájek, J. and Sidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.
- Halperin, M. and Ware, J. (1974). Early decision in a censored Wilcoxon two-sample test for accumulating survival data. J. Amer. Statist. Assoc. 69, 414-22.
- Johnson, R.A. and Mehrotra, K.G. (1972). Locally most powerful rank tests for the two-sample problem with censored data. Ann. Math. Statist. 43, 823-31.
- Koziol, J.A. and Petkau, A.J. (1978). Sequential testing of the equality of two survival distributions using the modified Savage statistic. Biometrika 65, 615-23.
- Lesser, M.L. (1979). Two-sample nonparametric sequential life testing using the progressively censored Savage statistic. Unpublished Ph.D. dissertation, Rutgers University.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports 50, 163-170.
- Mantel, N. and Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22, 719-48.
- Martin, A.R. (1946). The use of mice in the examination of drugs for chemotherapeutic activity against mycobacterium tuberculosis. J. Pathol. Bacteriol. 58, 580-5.
- Muenz, L.R., Green, S.B. and Byar, D.P. (1977). Applications of the Mantel-Haenszel statistic to the comparison of survival distributions. Biometrics 33, 617-26.

Mosteller, F. and Tukey, J.W. (1977). Data Analysis and Regression.
Addison-Wesley, Reading, Pa.

Peto, R., Pike, M.C., Armitage, P., Breslow, N.E., Cox, D.R., Howard, S.V.,
Mantel, N., McPherson, K., Peto, J., and Smith, P.G. (1976,77).
Design and analysis of randomized clinical trials requiring prolonged
observation of each patient I, II. Br. J. Cancer 34, 585-612;
35, 1-39.

Robinson, J. (1978). An asymptotic expansion for samples from a finite
population. Ann. Statist. 6, 1005-11.

Savage, I.R. (1956). Contributions to the theory of rank order statistics:
the two-sample case. Ann. Math. Statist. 27, 590-615.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biometrics
Bull. 1, 80-83.