DETERMINATION OF P-VALUES FOR A K-SAMPLE EXTENSION OF THE KOLMOGOROV-SMIRNOV PROCEDURE

by

Erica H. Brittain

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1472

November 1984

DETERMINATION OF P-VALUES FOR A K-SAMPLE EXTENSION OF THE KOLMOGOROV-SMIRNOV PROCEDURE

by

Erica H. Brittain

A Dissertation submitted to the faculty of The University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill

1984

Approved by:

Clarence E. Davis
Adviser

Reader

Dana Quade
Reader

A B S T R A C T

ERICA HYDE BRITTAIN. Determination of p-values for a k-sample extension of the Kolmogorov-Smirnov procedure (Under the direction of C.E. DAVIS and T.R. FLEMING.) A natural statistic for a k-sample extension of the Kolmogorov-Smirnov procedure is the maximum of the two-sample Kolmogorov-Smirnov statistics obtained from each pair of samples. This statistic was discussed by Birnbaum and Hall (1960), among others. However, its distribution in general has remained unknown. This paper first determines approximate p-values for the equal sample size case; the approach is later extended to unequal sample sizes. The strategy is to regard the p-value for the value T as the probability of a union of events, where for each sample pair the corresponding event is the sample space on which the two-sample statistic equals or exceeds the value T. This probability equals k sums of intersections of these events. P-values under .10 can be very well approximated by the first three sums. It is found that the k-sample p-value (for k up to ten samples) for T is approximately a simple curve function of the two-sample p-value for T; a curve for each k is presented. The magnitude of error introduced by the approximations is investigated directly and also by simulations and found to be negligible except when there are eight or more samples and the p-value is greater than .05. We discuss why the k-sample p-value is a function of the two-sample p-value, even for small samples. An example of the method is presented, which compares the distribution of triglycerides and of cholesterol in the four race-sex groups for children 10-11 years old.

ACKNOWLEDGEMENTS

First of all, I am very grateful for advice, suggestions, and encouragement I received from the members of my committee: Ed Davis, Tom Fleming, Dana Quade, Kant Bangdiwala, and Al Tyroler. I also wish to thank Dr. Charles Glueck who granted me permission to use the data which appear in Chapter Seven.

I particularly wish to thank Tom Fleming who agreed to advise me under unorthodox circumstances. His kindness in doing so will always be remembered. Furthermore, he devoted much time and provided much insight. I also am indebted to Ed Davis, who helped with all the special arrangements and whose advice always proved sound. Ed Davis and Tom Fleming, along with Jim Grizzle and Mike O'Fallon, were all instrumental in setting up an arrangement where I could work on my dissertation at the Mayo Clinic, thus allowing me to marry my fiance.

Finally, I wish to thank my family for love and encouragement. My extra special thanks go to my husband, John Hyde, who provided me with everlasting loving support and was a continual source of helpful suggestions and enthusiasm.

TABLE OF CONTENTS

C H A P T E R   O N E

LITERATURE REVIEW

KOLMOGOROV

Kolmogorov (1933) proposed a goodness of fit test which compares the empirical cumulative distribution function (cdf) of a sample to a specified cumulative distribution function. His procedure was designed to test H_o: $F(x) = U(x)$, where $F(x)$ is the cdf of the population of interest and $U(x)$ is the hypothesized cdf. He devised the statistic

$$K_n = n^{1/2} \sup_{-\infty < x < \infty} |F_n(x) - U(x)|,$$

where F_n is the empirical cdf of a sample from F. He proved that the distribution of K_n is independent of $F(x)$ if H_o is true and $F(x)$ is continuous; thus the test is distribution free. Furthermore, he derived the limiting distribution:

$$\lim_{n \to \infty} P(K_n < x) = \lim W_n(x) = W(x)$$

$$= \sum_{j=-\infty}^{\infty} (-1)^j \exp(-2j^2x^2)$$

He also noted that if $F(x)$ is not continuous then $P(K_n<x) \geq W_n(x)$; therefore the test would be conservative if $F(x)$ is not continuous.

SMIRNOV

Smirnov (1939) expanded Kolmogorov's work to the two sample setting; he also confirmed Kolmogorov's limiting distribution using a different approach. Together, these procedures are frequently referred to as the one- and two-sample Kolmogorov-Smirnov tests. Smirnov provided a distribution free test of H_0: $F(x) = G(x)$, where $F(x)$ and $G(x)$ represent unspecified cdf's from two populations. His two sided statistic is as follows:

$$D_{mn} = [mn/(m+n)]^{1/2} \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|$$

where n is the number of observations in the sample from F and m is the number of observations in the sample from G. Smirnov showed that D_{mn} has the same limiting distribution as K_n. Smirnov also developed several one-sided tests:

$$K_n^+ = n^{1/2} \sup_{-\infty<x<\infty} (F_n(x) - F(x))$$

$$D_{mn}^+ = [mn/(m+n)]^{1/2} \sup_{-\infty<x<\infty} (F_n(x) - G_m(x))$$

$$D_{mn}^- = [mn/(m+n)]^{1/2} \sup_{-\infty<x<\infty} (G_m(x) - F_n(x))$$

Smirnov demonstrated:

$$\lim_{n\to\infty} P(K_n^+ < x) = \lim_{n,m\to\infty} P(D_{mn}^+ < x) = 1 - \exp(-2x^2)$$

$$\lim_{m,n\to\infty} P(D_{mn}^+ < x;\ D_{mn}^- < y) =$$

$$1+\sum_{j=1}^{\infty}\left(2\exp(-j^2(x+y)^2)-\exp(-2(jx+(j-1)y)^2)-\exp(-2(jy+(j-1)x)^2)\right)$$

In a later paper, Smirnov (1948) published tables of the limiting distribution to make these results more accessible.

Numerous authors have pointed out that these asymptotic distributions derived by Kolmogorov and Smirnov are not very accurate for even a moderately large n, 50 say. Some have tried to improve on Smirnov's formula to provide more accurate p-values for small and moderate n. For example, Kim (1969) developed an alternate formula which yields p-values for Kolmogorov-Smirnov statistics which more closely match the exact p-values than the asymptotic formula, when $n < 100$. To devise this formula, he used the moments to determine that a certain function of the statistic has a normal distribution.

SMALL SAMPLE DISTRIBUTIONS FOR THE KOLMOGOROV-SMIRNOV TEST

Finding small sample distributions for the two-sample test has posed considerable difficulty. Gnedenko and Korolyuk (1951) derived exact distributions for D_{mn} when n=m, using random walk methodology. Their proof was based on the observation that $P(D_{nn} \geq h/n)$ equals the probability that the largest distance from the origin of a random walk on a line is at least h. The random walk starts at the origin and the path takes 2n steps, n to the right and n to the left; all possible path orders are equally likely. The probability that a path deviates from the origin by at least h is computed by the reflection principle. Gnedenko and Korolyuk demonstrated:

$$P(D_{nn} \geq h/n) = 2(C_{2n,n})^{-1} \sum_{i=1}^{n/h} (-1)^{i+1}(C_{2n,n-ih}),$$

note: $(C_{a,b})$ denotes the number of a choose b combinations

Distributions of the one-sided statistic have been derived by Korolyuk (1955) and by Hodges (1957). Korolyuk's derivation applies only to the case where one sample size is a multiple of the other, while Hodges' method applies only when the sample sizes differ by one. Steck (1969) stated that formulas for the distribution of two-sided statistics are extremely complex and not well suited for computation.

Efficient geometric algorithms for computing exact p-values have been developed. For example, one approach introduced by Hodges (1957) systematically counts the number of paths whose coordinates remain inside a region of non-significance. Each path starts at (0,0); its first coordinate increases by one after each observation in Sample 1, and its second coordinate increases by one after each observation in Sample 2. The path ultimately ends up at (n,m). The recursive nature of this counting process renders this a simple task. The final count is compared to the total number of possible paths to yield the exact p-value. Tables of p-values achieved by similar means have been published by Massey (1951,1952), Birnbaum and Hall (1960), and Kim and Jennrich (1970). Furthermore, one can easily determine the p-value for a given statistic for any combination of n and m using a brief computer routine.

K SAMPLES

Numerous k-sample analogs of the Kolmogorov-Smirnov procedure have been proposed. There is no single obvious k-sample generalization; none which is uniquely theoretically correct. In view of the difficulties in determining the distribution of the two-sample statistic, it is not surprising that the distributions of the various k-sample statistics have proven to be most elusive. Almost all work has been confined to the realm of equal sample sizes. The distributions of the following statistics have been investigated: T_1 by David (1958), T_2 by Kiefer (1959), T_3 by Birnbaum and Hall (1960), T_4 by Conover (1965), T_5 by Conover (1967), and T_6 by Wallenstein (1980).

$$T_1 = \max\{\sup_x(F_2(x)-F_1(x)), \sup_x(F_3(x)-F_2(x)), \sup_x(F_1(x)-F_3(x))\}$$

$$T_2 = \sup_x \sum_{i=1}^{k} n_i\ [F_i(x) - F(x)]^2,$$

where $F(x)$ is the pooled cdf,

$$T_3 = \sup_{x,i,j} |F_i(x) - F_j(x)|$$

$$T_4 = \sup_x [F^{(1)}(x) - F^{(k)}(x)]$$

where $F^{(j)}(x)$ represents the cdf of the sample of rank j

$$T_5 = \sup_{x,i} [F_i(x) - F_{i+1}(x)]$$

$$T_6 = \sum_{i=1}^{k-1} \sup_x [F_i(x) - F_{i+1}(x)]$$

T_1 through T_4 are two-sided procedures and T_5 and T_6 are both one-sided procedures. T_1 only applies to the case where $k=3$.

David (1958) sought a three-sample statistic which was amenable to an extension of Gnedenko and Korolyuk's geometric work. He asserted that an obvious k-sample analog T_3 was not such a statistic. He proposed T_1 as an alternative to T_3 since its distribution could be obtained by this approach. He generalized Gnedenko and Korolyuk's line to a plane. His path jumped one unit in direction 0 after each observation from Sample 1, a unit step in direction $(2/3)\pi$ after each observation from Sample 2, and a unit step in direction $(4/3)\pi$ after each observation from Sample 3. By considering all possible paths, David arrived at the following result, under the restriction of equal sample sizes:

$$P(T_1 \geq h/n) = 3 \sum_{i=1}^{n/h} \sum_{j(i)} \pm (n!)^3/(n-ih)!(n+jh)!(n+ih-jh)!$$

where set $j(i)$ consists of the integers $(2-i,3-i,5-i,6-i,8-i,9-i,11-i,12-i,\ldots 2i)$, and $\pm$ indicates that the successive terms have alternating signs, starting with + for $(2-i)$.

He also derived a large sample distribution for this statistic:

$$\lim_{n\to\infty} P(n^{1/2}T_1 \geq R) = 3 \sum_{i=1}^{\infty} \sum_{j(i)} \pm \exp(-R^2(i^2 + j^2 - ij)),$$

David's work was very clever, but the statistic itself is somewhat unsatisfactory. Consider the following situation: Population 1 has the smallest observations, Population 3 is in the middle, and Population 2 has the largest. This statistic is not powerful for detecting that kind of deviation between Populations 1 and 2. If, however, the labels for 2 and 3 were switched, the statistic could now detect that same difference. This is certainly an undesirable property.

Kiefer (1959) furnished the limiting distribution for T_2 for general k. It does allow for sample sizes which are not equal, but requires that each sample size be large. Using stochastic processes he determined:

$$\lim P\,[T_2 \leq a] = A_{k-1}(a)$$

$$\text{where } A_h(a) = P\{\max_{0\leq t\leq 1} \sum_{i=1}^{h} [Y_i(t)]^2 \leq a\}$$

is tabled for h up to 5 in his paper

($Y_i(t)$ are independent "tied down Wiener processes")

He noted that finding the distribution of T_3 using this methodology would be extremely difficult.

Birnbaum and Hall (1960) tabulated exact p-values for k=3; equal $n \leq 40$ for statistic T_3 using a three dimensional path counting recursive algorithm, a simple extension of Hodges' work. They pointed out that for k>3, the computations are theoretically no more complex, but become prohibitive due to the excessive demand on the computer memory. They suggested the use of the following:

for $k \geq 3$: $P(T_3(k) \leq r) \geq 1 - (C_{k,2})P(T_3(2) > r)$

for $k \geq 4$: $P(T_3(k) \leq r) \geq 1 - (C_{k,3})P(T_3(3) > r)$

Note that the first inequality is simply Bonferroni's Inequality in disguise. Birnbaum and Hall indicated that these could be regarded as approximate equalities at conventional significance levels. However, they conceded that these would result in conservative tests.

Taylor and Becker (1982) extended Birnbaum and Hall's work to include tables for k=4 for equal n from one to ten. They also tabulated the exact p-values for k=3 for some unequal sample sizes.

One can combine the tabulated results of Birnbaum and Hall with those of Taylor and Becker to see the highly conservative nature of these approximate equalities proposed by Birnbaum and Hall. For instance, when n=10, $P(T_3(3) \leq .7) = .99411$, so the inequality tells us that $P(T_3(4) \leq .7) \geq 1-4(1-.99411) = .976$. But the exact probability is .989 according to Taylor and Becker's calculations.

Gardner, Pinder and Wood (1979) conducted a simulation study to obtain estimated quantiles for T_3. They performed 5000 replications for each of various n,k combinations. This produced estimates of p-values, such that an estimated p-value of .05 had a 95% confidence interval of (.044,.056). The paper, however, only reported the statistics at various quantiles rather than the actual p-values.

Conover (1965) concocted a two-sided k-sample test for equal sample sizes. He designated the sample with largest observation of all as having "rank" k. The sample whose largest observation is smaller than the largest observation of any other sample has "rank" one. The statistic T_4 is essentially the regular two-sample statistic comparing the sample with rank k to the sample with rank one. Conover indicated this statistic could be used to pick up a difference in the location parameter. He determined an exact expression for $P(T_4 \leq c/n)$ which has a complex but closed algebraic form. He also discovered the asymptotic distribution of the statistic and realized that it was identical to the asymptotic distribution of the one-sided two-sample test. As an aside, Conover pointed out the danger in using the asymptotic result for the two-sample test, since it has the same form as that of T_4 regardless of the number of samples.

Conover's statistic is an intriguing one; however, its value is somewhat debatable. It seems inappropriate to ignore the information on all but two of the samples, especially since the criteria used for selecting the two samples depend on little information. It is very easy to imagine a set of circumstances wherein this statistic fails to pick up a meaningful difference between samples. It is overly sensitive to outliers.

Conover (1967) introduced another statistic T_5 which is for a one-sided testing situation; its use is restricted to the equal sample size case. His test is designed to detect the alternative $F_i(x) > F_j(x)$ for some $i < j$. Using an induction approach, he found an expression for the distribution of T_5 which is very complex but has closed algebraic form. He also ascertained the limiting distribution of the statistic.

Wolf and Naus (1973) verified Conover's (1967) distributional result using a different approach: the k-sample ballot problem. They provided a tabulation of critical values for Conover's test. By means of simulations, they investigated the power of the k-sample test relative to the ANOVA and Kruskal-Wallis procedures. They alleged that T_5 provided a powerful test against alternatives in which two adjacent populations had a large difference.

Wallenstein (1980) offered statistic T_6 which is based on the sum of (k-1) one-sided statistics, when the populations can be ordered a priori. Again, this statistic can only be applied to equal sample sizes. He derived the distribution of T_6 by extending the ballot box approach. He then tabulated the small sample distribution and also derived an asymptotic distribution. He claimed that T_6 is powerful against alternatives of trend and those of a single jump. After conducting some simulations, he found T_6 superior to T_5 in general. He recommended T_6 over Jonckheere's statistic when an investigator could propose an ordering, but was not very certain about its correctness.

CHAPTER TWO

DETERMINATION OF P-VALUES FOR A K-SAMPLE PROCEDURE

Our goal is to develop a method for determination of p-values using statistic T_3, which will from now on be referred to as T. T is a desirable statistic from several standpoints. T is a simple and obvious extension of the two-sample test and it is two-sided, which is probably the more common situation for the k-sample setting. Its largest drawback is its lack of recognition of differences in sample sizes; it probably is only appropriate for those settings in which the sample sizes are equal or at least reasonably well balanced.

Our method seeks to elucidate the approximate relationship between two-sample p-values and k-sample p-values. Once this relationship is established, the task of determining the k-sample p-values is done. Consider the statistic T again:

$$T = \sup_{x,i,j} |F_i(x) - F_j(x)|$$

In practice, one could generate the two-sample p-values for each of the "k choose 2" pairings of samples. If any of these pairings result in a test which is significant at level α_2, then the entire k-sample procedure is significant at some value α_k. Or, equivalently, one could find the smallest two-sample p-value (which corresponds to the largest two-sample statistic) and therefore determine the procedure-wise p-value, if the correspondence between the two-sample and the k-sample p-value were known.

This approximate correspondence is determined by use of elementary probability laws as well as some approximations. The end product is a curve for each k which plots two-sample p-values against k-sample p-values.

CORRESPONDENCE BETWEEN TWO- AND THREE-SAMPLE P-VALUES

Examination of the tables presented in the Birnbaum and Hall (1960) paper generated an interesting and valuable discovery. There appears to be an approximate one-to-one correspondence between two- and three-sample p-values _regardless of the sample size, n_. For example, consider the following p-values derived from the Birnbaum-Hall tables:

n=11, T=.636		n=18, T=.500	
two-sample	three-sample	two-sample	three-sample
.02074	.05504	.02075	.05508

In other words, if the two-sample p-value for a given statistic is about .0207, then the three sample p-value associated with the same statistic is about .055. Thus, if one determined the p-value for each of the three pairings of samples and found the smallest p-value to be .0207, then the three-sample p-value for that experiment would be about .055, regardless of n.

A series of two-sample p-values (for sample sizes from 4 to 40) are plotted against their corresponding three-sample p-values in Figure 1. The points lie very close to a simple hand drawn curve, thus demonstrating that the three-sample p-value is simply a function of the two-sample p-value at least in an approximate sense. We should note that those pairs from small sample sizes, such as four, are very slightly off the curve. This suggests that the relationship exists regardless of n, once n is "large" and n becomes large very quickly. This observation generated the idea of reducing the k-sample problem to describing the relationship between the 2-sample and k-sample p-values.

BONFERRONI-TYPE APPROXIMATIONS TO K-SAMPLE P-VALUES

When k=3, we know that there exists a procedure-wise alpha level α which corresponds to the two-sample significance level γ. That is, there is the following relationship, where AB represents the two-sample T statistic comparing sample A to sample B:

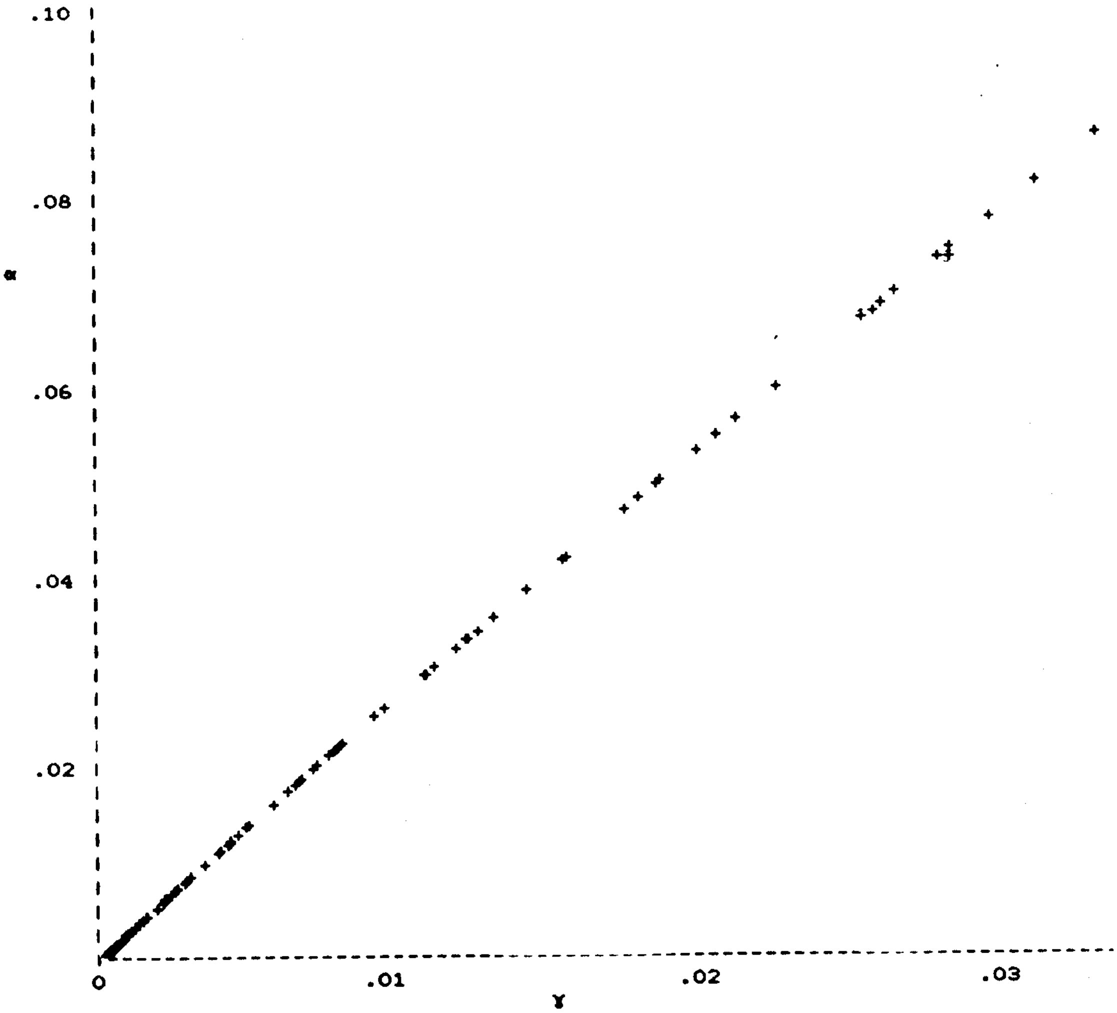

Figure 1: α vs. γ

$$\alpha = P(\max(AB, AC, BC) \text{ is significant at } \gamma)$$

$$\alpha = P(AB \text{ significant at } \gamma \text{ or } AC \text{ significant at } \gamma \text{ or } BC \text{ significant at } \gamma),$$

$$\begin{aligned}\alpha = {} & P(AB \text{ sig at } \gamma) + P(AC \text{ sig at } \gamma) + P(BC \text{ sig at } \gamma) \\ & -P(AB \text{ sig at } \gamma;\ AC \text{ sig at } \gamma) \\ & -P(AB \text{ sig at } \gamma;\ BC \text{ sig at } \gamma) \\ & -P(AC \text{ sig at } \gamma;\ BC \text{ sig at } \gamma) \\ & +P(AB \text{ sig at } \gamma;\ AC \text{ sig at } \gamma;\ BC \text{ sig at } \gamma).\end{aligned}$$

In general, α = P(at least one pair is significant at γ). This is the probability of a union of "k choose 2" events. Using the notation in Feller (1968), let S_1 denote the sum of the probabilities of each single event occurring; S_2 is the sum of the probabilities of intersections of each pair of events; and so forth. Thus, $\alpha = S_1 - S_2 + S_3 - S_4 + S_5 - \ldots\ldots \pm S_k$. As stated by Feller, Bonferroni's Inequality involves the following idea: if the terms S_1, S_2, S_3, ..., S_{r-1} are kept while the terms S_r, ..., S_k are omitted, then the exact value minus the approximation has the sign of the first dropped term and is smaller in absolute value.

Thus, α can be approximated by S_1 which is the familiar form of the "Bonferroni Inequality" and we know that this is an overestimate of α; the error is less than the value of

S_2. If the values of the S_i's were declining in value rapidly, one could get excellent precision by refining this approximation further. We could approximate α by $S_1 - S_2 + S_3$. This still would be an overestimate of α, but now we know the magnitude of the error is smaller than S_4.

Estimation of α by the first three components as suggested above is a realistic goal, since this limits the involved probability types to only eight. These eight probability types are as follows (for ease of notation we will use P(AB) as shorthand for P(AB significant at γ)). The pictures to the right of each probability graphically depict the involved relationships; two samples which are required to be significantly different from each other are connected by a line.

Type 1 P(AB)

```
A-B
```

Type 2 P(AB;AC)

```
  A
 / \
B   C
```

Type 3 P(AB;CD)

```
A-B   C-D
```

Type 4 P(AB;AC;AD)

```
  D
  |
  A
 / \
B   C
```

Type 5 P(AB;BC;CD) A-B-C-D

Type 6 P(AB;AC;BC)

```
  A
 / \
B — C
```

Type 7 P(AB;AC;DE)

```
  A
 / \     D-E
B   C
```

Type 8 P(AB;CD;EF) A-B C-D E-F

For example, if k=3, then α = 3P(Type 1) - 3P(Type 2) + P(Type 6), since under the null hypothesis P(AB) = P(AC) and so forth, when the sample sizes are all equal. It is straightforward to calculate the frequencies of each type for a value of k. In general, the frequencies are:

Type 1 $(C_{k,2})$

Type 2 $(C_{k-1,2})(k)$

Type 3 $(.5)(C_{k-2,2})(C_{k,2})$

Type 4 $(C_{k-1,3})(k)$

Type 5 $(12)(C_{k,4})$

Type 6 $(C_{k,3})$

Type 7 $(30)(C_{k,5})$

Type 8 $(15)(C_{k,6})$

Note $(C_{k,r}) = 0$ if $r>k$.

The remaining step is to determine the probability of each type. After this is accomplished, we can then estimate α by:

$$\sum_{i=1}^{8}(1-\delta_{i2})P(\text{Type } i)\text{Freq}(\text{Type } i),$$

where $\delta_{i2}=2$ when $i=2$ or 3 and δ_{i2} equals zero otherwise. The probability of each type is established over the next few pages. A number of approximations are invoked; it is hoped that the resulting errors will be too minimal to have any substantial impact.

PROBABILITIES OF EACH TYPE

Note: This procedure will not be as accurate for p-values greater than .10, as discussed later in this chapter. The following probability types are determined so that they are appropriate for small values of γ, such that the corresponding value of α is less than .10.

Probability of Type 1

P(AB is significant at γ) is simply γ by definition. Therefore, there is no error in the estimation of this probability.

Probability of Type 2

P(AB significant at γ and AC significant at γ) is more complicated than the preceding probability due to the lack of independence between AB and AC. Recall that the p-value for a three-sample statistic $\alpha = 3P(AB) - 3P(AB;AC) + P(AB;AC;BC)$. The last term, P(AB;AC;BC) can be regarded as zero and thus ignored, since the likelihood that "A" is significantly different from "B", "A" from "C", and "B" from "C" simultaneously under H_o for any conventional significance level γ is extremely remote. Furthermore, any error due to regarding this as zero is completely cancelled out by a compensating error in Type 6, as discussed later. Therefore α can be well approximated by $3(P(AB) - P(AB;AC))$. Using algebra, we observe that this probability type is estimated by $\gamma - \alpha/3$. We can now take known α and γ pairs from the Birnbaum-Hall tables and determine the associated probability.

Probability of Type 3

Calculation of P(AB;CD) is very simple since the sample pairings are independent. The probability of this type is exactly equal to $P(AB)P(CD) = \gamma^2$.

Probability of Type 4

The estimation of P(AB;AC;AD) is very difficult. It cannot be tackled directly; thus empirical arguments must suffice. The computer algorithm for obtaining p-values as used by Hodges and others, can be modified to determine the probability of any union of relevant events. P(AB or AC or AD) equals 3P(AB) - 3P(AB;AC) + P(AB;AC;AD). Thus, P(AB;AC;AD) = -3P(AB) + 3[2P(AB) -P(AB or AC)] + P(AB or AC or AD). The computer algorithm can calculate P(AB), P(AB or AC), and P(AB or AC or AD) for a specific sample size and critical value. Using the above relationship, we can find the corresponding P(AB;AC;AD). Due to computer storage limitations, P(AB;AC;AD) was only determined for sample sizes less than or equal to 17. We saw earlier that the relationship between γ and α was not as consistent for observations from small sample sizes, such as four. Consequently, we included only sample sizes greater than or equal to 8 for our empirical study. This provides for 29 observations based on sample sizes between 8 and 17 with γ values acceptably small. The appendix 2.1 at the end of this chapter displays a list of each observation and its corresponding probabilities including P(AB;AC;AD). Figure 2 illustrates the very strong and simple relationship between γ and P(AB;AB;AD) regardless of n for these 29 observations. All 29 pairs of γ and P(AB;AC;AD) were read into a power curve program (i.e. linear regression on the logarithms of both), yielding the following relationship: P(AB;AC;AD) = $.340665\gamma^{1.6348}$, with r^2=.99996. This function is also presented in Figure 2. The only reservation we may have about accepting this function as the probability type is that the relationship could possibly change with sample

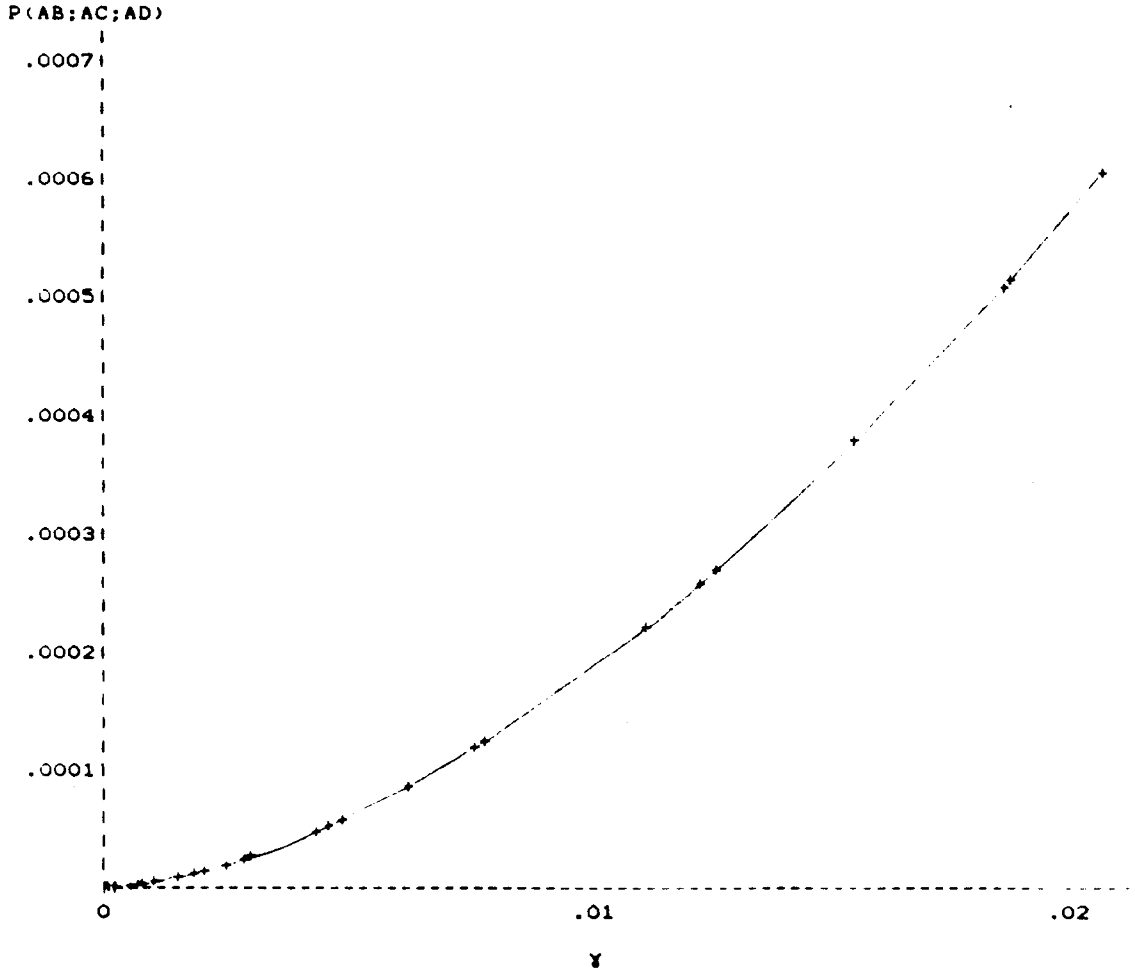

Figure 2: P(AB;AC;AD) vs. γ

sizes larger than 17. We explored this by calculating P(AB;AC;AD) for several statistic values for n=30 and for n=46. We observed that while the estimation is slightly less perfect than for small values of n, it still was excellent; the error was less than 2 per cent and served to very slightly underestimate the overall p-value. For example, for n=46, T=16/46, the exact probability is .000110, whereas the estimate is .000108. This suggests that the relationship stays virtually the same for n as large as 46.

Probability_of_Type_5

Again, an empirical approach was needed to find the probability of this type. Proceeding in a similar fashion as we did for Type 4, we can obtain P(AB;BC;CD) for various γ values drawn from sample sizes less than or equal to 17. Appendix 2.1 presents the specific probabilities for each observation. The plot of γ versus P(AB;BC;CD) is presented in Figure 3, illustrating a clear relationship once again, regardless of sample size (note: there are only 25 observations included in this analysis, since 4 values of P(AB;AC;BD) were too small to be accurately determined and were dropped). One can express P(AB;BC;CD) as P(BC|AB;CD)P(AB;CD) which equals P(BC|AB;CD)γ^2. We then observe that P(BC|AB;CD) varies from .263 and .315 in our sample of 25 and appears to have a slight linear relationship with γ. Regression of P(BC|AB;CD) on γ yields

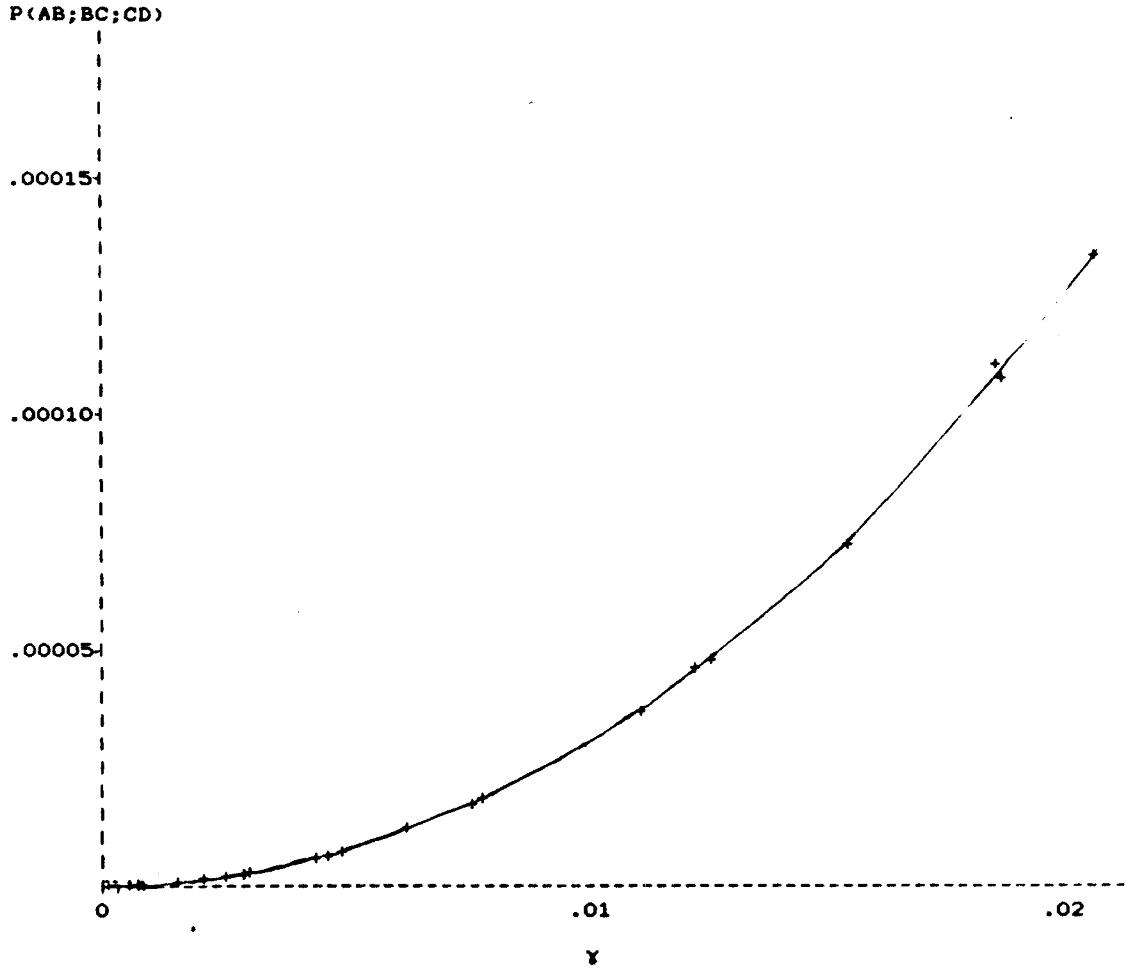

Figure 3: P(AB;BC;CD) vs. γ

the following relationship: P(BC|AB;CD) = .27732 + 1.4235γ. Note that this conditional probability ranges from .277 to .306 when γ ranges from .00 to .02. Thus, the probability type is estimated by $.27732\gamma^2 + 1.4234\gamma^3$, which is the function presented in Figure 3. This provided a better fit (r^2=.9996) than any other approach including the power curve regression between γ and P(AB;BC;CD) as done in the previous type. Again, we must have some concern about the possibility that this relationship changes as the sample size grows larger, but again we checked n=30 and n=46, and observed that the fit was as good as for the smaller n.

Probability of Type 6

As discussed in an earlier section, P(AB;AC;BC) must be extremely small as long as γ is below .05. As γ becomes smaller and smaller, as it must do for increasing values of k, then this probability approaches zero. The slight error caused by this approximation is completely erased by a compensating error in Type 2, as will be discussed later.

Probability of Type 7

This probability can be simply calculated based on already established probabilities. P(AB;AC;DE) = P(AB;AC)P(DE) = $(\gamma-\alpha/3)(\gamma)$.

Probability of Type 8

The P(AB;CD;EF) is exactly γ^3 due to independence.

The following table summarizes the above findings. The third column indicate whether the probability is exact or approximate. The last column indicates whether the probability is based on a theoretical or empirical result.

1 P(AB)	γ	E	T
2 P(AB;AC)	$\gamma-\alpha/3$	E	T
3 P(AB;CD)	γ^2	E	T
4 P(AB;AC:AD)	$.340665\gamma^{1.63484}$	A	E
5 P(AB;BC;CD)	$.27732\gamma^2 + 1.4235\gamma^3$	A	T;E
6 P(AB;AC;BC)	0	E	T
7 P(AB;AC;DE)	$(\gamma-\alpha/3)\gamma$	A	T
8 P(AB;CD;EF)	γ^3	E	T

Note: Types 2 and 6 are exact only when considered jointly.

ERROR IN THE ESTIMATION OF THE PROBABILITY TYPES

There are three types estimated approximately: 4,5 and 7. In addition, types 2 and 6 are exact only when considered jointly. In this section, we will discuss the error in these types.

Types_2_and_6

Let X=P(AB;AC), Y=P(AB;AC;BC), and α=P(AB or AC or BC). Recall:

$P(AB;AC)$ is estimated by $\gamma - \alpha/3$, and

$\alpha = 3\gamma - 3X + Y$.

Thus, the probability of Type 2, $P(AB;AC)$, is estimated by $\gamma - [3\gamma - 3X + Y]/3$. This quantity equals $X - Y/3$.

The frequency of Type 2 is $3(C_{k,3})$ and the frequency of Type 6 is $(C_{k,3})$. Since Type 2 is a member of term S_2 it is subtracted whereas Type 6 is added since it is a component of term S_3. Thus the true joint contribution of these two types is $(C_{k,3})[Y - 3X]$.

The joint contribution of the estimates of these two types is $(C_{k,3})[0 - 3(X - Y/3)]$, which equals the above quantity which represented the true joint contribution. In other words, the estimated type 2 probability by itself serves as a joint contribution of both types. Therefore, these two types are estimated without error.

Types 4 and 5

These two probability types are based on fitted equations. As mentioned above, while both have extremely stable relationships with γ for the sample sizes up to 17, we must assume that the relationship will remain the same as sample sizes increase. When we checked the relationship at moderate sample sizes, we observed a slight increase in error for Type 4 and none for Type 5. However, in both cases, the observed error was less than 2 per cent at the worst. To evaluate the effect of the potential error on the procedure-wise p-value, let us consider two situations. The first represents a likely circumstance, whereas the second represents the very worst case scenario. The first two columns in the following table present the total error produced when a 2 per cent error exists in the estimation of the type probability, and the procedure-wise p-value is about .05. As we can see, these errors are quite trivial. The second two columns represent an absolute worst case scenario. These values are based on an assumption of a 5 per cent error (far worse than observed), when the procedure-wise p-value is about .10. Under this circumstance, there would be some error in the fourth decimal place.

Error in Procedure-wise P-value

k	p-value=.05 2% error in Type 4	p-value=.05 2% error in Type 5	p-value=.10 2% error in Type 4	p-value=.10 2% error in Type 5
4	.00001	.00001	.00012	.00008
6	.00005	.00002	.00048	.00022
8	.00009	.00002	.00077	.00028
10	.00012	.00003	.00102	.00030

It should be noted that the earlier suggestion indicated that in moderate sample sizes, the error in Type 4 would serve to underestimate the p-values, whereas the error in Type 5 may be in the other direction. As will be discussed in the upcoming chapter on the fourth S-term, the magnitude of the error at alpha equal to .10 that is due to the dropped S-terms which causes an over-estimate, is considerably larger than the above error in the worst case setting. Thus, even if this worst case occurred, it is likely to only ameliorate the overall error.

In conclusion, the error in the p-value resulting from the slight estimation errors in Types 4 and 5 is likely to be negligible, and in some cases may even slightly compensate for the overestimation due to the dropped S-terms.

Type 7

Again, let X=P(AB;AC) and Y=P(AB;AC;BC). As shown above, P(AB;AC) is estimated by: X - Y/3. Thus, P(AB;AC;DE) equals γX, but is estimated by γX -γY/3. Consequently, the probability of Type 7 is underestimated by γY/3. Moreover, the overall resulting error in the estimation of the p-value is the product of the frequency of Type 7 and γY/3. Consider the following examples:

					P(AB;AC;DE)		
k	Freq	γ	n	T	exact	estimate	error
5	30	.01234	10	.700	.000014313	.000014267	.0000015
7	630	.00629	9	.777	.000002851	.000002845	.0000040
8	1680	.00490	14	.643	.000001553	.000001548	.0000079
10	7560	.00301	16	.625	.000000484	.000000483	.0000073

The above displays the worst possible case for each respective k. The error will decrease as γ decreases and for each k; these examples represent the largest γ will be so that the overall p-value does not exceed .10. Thus, these examples have the largest possible error for their respective numbers of samples. This suggests that the entire underestimation due to the error in Type 7 is less than .00001.

CALCULATION OF K-SAMPLE P-VALUE CURVES

We have now established frequencies and estimated probabilites for each type. We can now estimate the k-sample α corresponding to the two-sample γ. Then, if any two-sample pairs are significant at γ, the whole procedure is significant at α.

By using already known γ,α pairs drawn from the Birnbaum-Hall tables, α_3 through α_k can be computed for each pair. However, for increased accuracy over the numbers with 6 decimal places presented in the Birnbaum-Hall tables, we recomputed these pairs to 10 decimal places. (As an important aside: we discovered that the Birnbaum-Hall tables are in error for the p-values for three samples of size 24. The correct probabilities are given in Appendix 2.2.). Then for each k, the γ,α_k points are plotted. The points fall on an easily discernible curve, as drawn on Figure 4, for k=3 through 10. Note that near the origin, each curve is a line with slope of $(C_{k,2})$, which corresponds

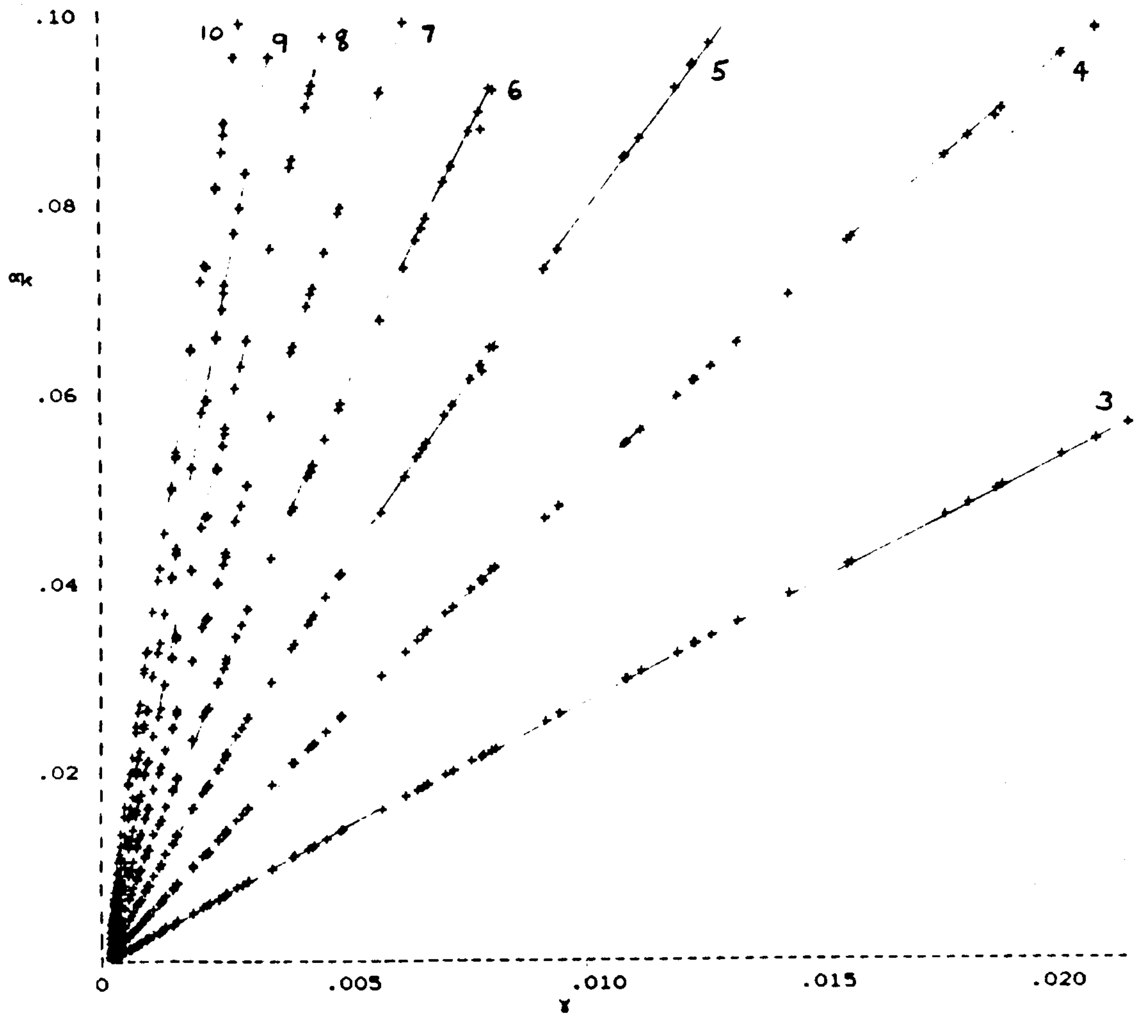

Figure 4: α_k vs. γ

to the Bonferroni inequality. Note that p-values greater than .10 do not appear on this graph. This is because the method developed here gets increasingly inaccurate as the p-values increase, especially for the higher values of k. Likewise, p-values for k greater than 10 can be obtained by this approach, but they will be increasingly inaccurate. This is because the S terms which are dropped are not so negligible when either γ grows large or k grows large. The dropped S terms grow with γ because the probability of several simultaneously significant tests becomes much less remote. The dropped S terms grow with k because the number of added probabilities within each term gets very large. In addition, the estimation of some of the probability types is flawed as γ becomes large.

USE OF THE CURVES

The algorithm for using the k-sample curves is best illustrated by means of an example. For example, say one has data from six samples, and has found that the smallest p-value yielded from the "6 choose 2" or 15 sample pair tests is .006. Then .006 is located on the γ axis of Figure 4; the six sample curve intersects the line γ=.006 at α_6 = .070. Therefore, the procedure-wise p-value is .070.

EQUATIONS

Near the origin, each curve appears to be a line of slope $(C_k,2)$. To fit that part of the curve which deviates from this line, the (γ,α) points were transformed to $(\gamma, (C_k,2)\gamma - \alpha)$ points. The resulting points suggested power curves, i.e. $(C_k,2)\gamma - \alpha = \delta\gamma^{\beta}$ for each k. The points were entered into a power curve regression program and yielded equations for each k with very good fit. (r^2 = .9999). The equations were then transformed so that α was now a function of γ, and r^2 was then computed for this final curve. In the calculation, all pairs whose alpha-level exceeded .10 were removed from the analysis, since the curves are not reliable after this point. The results were:

Equations

k	Equation	r^2 deviant	r^2 final
3	$\alpha=3\gamma-1.5735\gamma^{1.3916}$	.99997	.9999997
4	$\alpha=6\gamma-5.3761\gamma^{1.3755}$	.99997	.9999999
5	$\alpha=10\gamma-11.4256\gamma^{1.3594}$	.99995	.9999988
6	$\alpha=15\gamma-19.3440\gamma^{1.3431}$	.99991	.9999999
7	$\alpha=21\gamma-28.4718\gamma^{1.3263}$	.99986	.9999656
8	$\alpha=28\gamma-37.5653\gamma^{1.3073}$	.99976	.9999419

9	$\alpha=36\gamma-47.4433\gamma^{1.2913}$	.99969	.9999215
10	$\alpha=45\gamma-54.3065\gamma^{1.2693}$	.99951	.9999000

These functions are illustrated in Figure 4. They have extraordinarily good fit and can be used to quickly calculate the procedure-wise p-value. As an alternative, one could calculate the p-value by actually plugging γ into all the probability types, thus bypassing this last approximating step; however, this would probably not be a worthwhile effort as opposed to the quick curve calculation.

UPCOMING CHAPTERS:

This second chapter has developed a procedure for determination of k sample p-values for the Kolmogorov-Smirnov test. The remaining chapters will cover the following topics:

3: Examination of the magnitude of the error due to omission of all but the first three S terms

4: Assessment of overall accuracy of the curves

5: Extension of the procedure to unequal sample sizes

6: Exploration of the underlying theory of the relationship between the two- and k-sample p-values

7: An example of use of the procedure

8: Discussion of future related work

APPENDIX 2.1

EXACT PROBABILITIES USED FOR TYPES 4 AND 5

For each observation, information appears in the following order:

1. P(AB) 2. P(AB or AC) 3. P(AB or AC or AD)

4. P(AB or BC or CD) 5. P(AB;AC) 6. P(AB;AC;AD)

7. P(AB;BC;CD) 8. Sample Size 9. Critical Value

0186480186	.0352065338	.0515265099	.0501829685	.0020895034
.0005074229	.0001092095	n=8	D=6/8	
.0024864025	.0048477797	.0072047955	.0071026002	.0001250253
.0000184686	.0000018208	n=8	D=7/8	
.0001554002	.0003080808	.0004607449	.0004582319	.0000027196
.0000001901	.0000000076	n=8	D=8/8	
.0062937063	.0121343775	.0179470804	.0176068653	.0004530351
.0000848517	.0000116424	n=9	D=7/9	
.0007404360	.0014577974	.0021747666	.0021546343	.0000230746
.0000025501	.0000001560	n=9	D=8/9	
.0123406006	.0235211861	.0345950619	.0337983715	.0011600151
.0002566150	.0000455807	n=9	D=7/10	
.0020567667	.0040186512	.0059774945	.0058993116	.0000948822
.0000136581	.0000011891	n=10	D=8/10	
.0002165018	.0004288305	.0006411254	.0006373308	.0000041731
.0000003447	.0000000131	n=10	D=9/10	
.0207390648	.0390731033	.0571089484	.0556060457	.0024050263
.0006039302	.0001319154	n=11	D=7/11	
.0043661189	.0084619637	.0125441572	.0123343992	.0002702741
.0000468648	.0000054117	n=11	D=8/11	
.0006549178	.0012905409	.0019258519	.0019089987	.0000192947
.0000021294	.0000001168	n=11	D=9/11	
.0000623731	.0001240041	.0001856323	.0001849388	.0000007421
.0000000458	.0000000011	n=11	D=10/11	

.0078590141	.0151040655	.0223051805	.0218579243	.0006139627
.0001227701	.0000178277	n=12	D=8/12	
.0014969551	.0029331520	.0043677207	.0043168207	.0000607582
.0000082300	.0000006127	n=12	D=9/12	
.0002041302	.0004044333	.0006047057	.0006012323	.0000038271
.0000003230	.0000000110	n=12	D=10/12	
.0126492702	.0241026321	.0354430753	.0346282289	.0011959083
.0002681432	.0000470853	n=13	D=8/13	
.0028748341	.0055990793	.0083173454	.0081966148	.0001505889
.0000238792	.0000022856	n=13	D=9/13	
.0004999712	.0009866981	.0014732415	.0014615785	.0000132443
.0000013978	.0000000665	n=13	D=10/13	
.0187822497	.0354797134	.0519305236	.0506063441	.0020847860
.0005139530	.0001061194	n=14	D=8/14	
.0048997173	.0094827592	.0140485669	.0138060971	.0003166754
.0000569714	.0000067730	n=14	D=9/14	
.0010207744	.0020058867	.0029902388	.0029597808	.0000356621
.0000044439	*	n=14	D=10/14	
.0076558083	.0147207899	.0217439167	.0213131469	.0005908267
.0001182021	.0000167566	n=15	D=9/15	
.0018373940	.0035940671	.0053481731	.0052815931	.0000807209
.0000115738	*	n=15	D=10/15	
.0111993539	.0213920604	.0314957061	.0307985764	.0010066474
.0002204569	.0000363647	n=16	D=9/16	
.0030152107	.0058695717	.0087172084	.0085890584	.0001608497
.0000259754	.0000023672	n=16	D=10/16	
.0006700468	.0013201978	.0019700854	.0019527404	.0000198958
.0000022874	*	n=16	D=11/16	
.0155606407	.0295236729	.0433159203	.0422673604	.0015976085
.0003782638	.0000713488	n=17	D=9/17	
.0046105602	.0089303315	.0132347713	.0130112513	.0002907889
.0000519374	.0000059258	n=17	D=10/17	
.0011526401	.0022630325	.0033724626	.0033366406	.0000422477
.0000054634	*	n=17	D=11/17	

*A precise estimate of P(AB;AC;BD) is not available.

APPENDIX 2.2

PROBABILITIES FOR THREE SAMPLES OF SIZE 24

Corrections to Birnbaum-Hall Tables

Critical Value	p-value	Notation Used in Tables: nr	Notation Used in Tables: $P(D(n,n,n) \leq r)$
10/24	.077758	9	.922242
11/24	.032590	10	.967410
12/24	.012165	11	.987835
13/24	.004037	12	.995963
14/24	.001187	13	.998813
15/24	.000307	14	.999693

CHAPTER THREE

ERROR DUE TO THE OMISSION OF FINAL S TERMS IN THE BONFERRONI EXPANSION

INTRODUCTION

A source of error in the estimation of the p-value curves will be explored in this chapter. It is that error due to the dropped S terms in the Bonferroni expansion. S_4 is the critical S term, since the Bonferroni Inequality tells us that it serves as an upper bound on the error resulting from the omission of it and all subsequent S terms. We can observe in our estimation of the first three S terms that they decline steeply, that is S_1 is much larger than S_2 and S_2 is much larger than S_3. Consequently, we speculate that S_4 (and thus the error) is likely to be much smaller than S_3, which rises with increasing k and α. Therefore we expect that S_4 is very small except when k is near 10 and α approaches .10. We will now determine an actual estimate for S_4 in precisely the same manner used to obtain the estimates of the previous S terms.

S_4 PROBABILITY TYPES

Estimation of S_3 involved only five probability types. Estimation of S_4 requires determination of eleven probability types. These types and their graphical representations are as follows:

Type 4.1 P(AB;AC;AD;BC)

```
    D
    |
    A
   / \
  B - C
```

Type 4.2 P(AB;AC;BD;CD)

```
  A - B
  |   |
  C - D
```

Type 4.3 P(AB;AC;AD;BE)

```
        D
        |
        A
       / \
  E - B   C
```

Type 4.4 P(AB;AC;AD;AE)

```
  D   E
   \ /
    A
   / \
  B   C
```

Type 4.5 P(AB;AC;BC;DE)

```
   A     D-E
  / \
 B - C
```

Type 4.6 P(AB;BC;CD;DE)

```
A-B-C-D-E
```

Type 4.7 P(AB;AC;AD;EF)

```
   D
   |
   A      E-F
  / \
 B   C
```

Type 4.8 P(AB;AC;DE;DF)

```
   A       D
  / \     / \
 B   C   E   F
```

Type 4.9 P(AB;AC;BD;EF)

```
C-A-B-D   E-F
```

Type 4.10 P(AB;AC;DE;FG)

```
   A      D-E     F-G
  / \
 B   C
```

Type 4.11 P(AB;CD;EF;GH)

```
A-B   C-D   E-F   G-H
```

The frequencies for each probability type are as follows:

Type 4.1	P(AB;AC;AD;BC)	$12(C_{k,4})$
Type 4.2	P(AB;AC;BD;CD)	$3(C_{k,4})$
Type 4.3	P(AB;AC;AD;BE)	$60(C_{k,5})$
Type 4.4	P(AB;AC;AD;AE)	$5(C_{k,5})$
Type 4.5	P(AB;AC;BC;DE)	$10(C_{k,5})$

Type 4.6	P(AB;BC;CD;DE)	$60(C_k,_5)$
Type 4.7	P(AB;AC;AD;EF)	$60(C_k,_6)$
Type 4.8	P(AB;AC;DE;DF)	$90(C_k,_6)$
Type 4.9	P(AB;AC;BD;EF)	$180(C_k,_6)$
Type 4.10	P(AB;AC;DE;FG)	$315(C_k,_7)$
Type 4.11	P(AB;CD;EF;GH)	$105(C_k,_8)$

PROBABILITY OF EACH TYPE IN S_4

Some of these probabilities factor into quantities which have already been established. However, for five of these types this is not possible. In these cases, we had to turn to the kind of empirical estimation used for several types in S_3. However, three of these cases involve five samples, so computer memory limitations only allowed exact determination of the probabilities for sample sizes less than or equal to eight. As a result, the final regression curve formulas are each based on only the 6 or 7 points which correspond to γ in the appropriate range. Fortunately, as seen in earlier estimations, these points consistently fall extremely close to a simple power curve. Given our experience in the earlier calculations, these power curve expressions are likely to be reasonably reliable when we extrapolate out for sample sizes greater than eight. The remaining two types only involve four samples and thus were based on a subset of those observations used to determine types 4 and 5 in S_3 (sample sizes range from 8 to 15).

Probability of Type 4.1

P(AB;AC;AD;BC) = P(AD|AB;AC;BC)P(AB;AC;BC).

We claimed earlier that P(AB;AC;BC) can be closely approximated by zero. Thus, this new probability could reasonably be approximated by zero. However, since the maximum value this probability can take on is unknown, an estimated probability was sought. The exact probability was determined for each of fifteen observations from the data in Appendix 2.1. These probabilities and their corresponding γs were entered into a power curve regression program. The resulting formula for this probability was $.130518\gamma^{2.56289}$ with $r^2=.9964$.

Probability of Type 4.2

P(AB;AC;BD;CD) unfortunately does not factor into previously established quantities. Thus, we obtained an empirical estimate of this type as a function of γ. The estimate of the probability type had adequate fit ($R^2=.9991$), however, not as good as that of some of the other types determined in this same manner. Our regression analysis of 21 observations from Appendix 2.1 produced this approximation:

$$.1475403\gamma^{2.0086297}.$$

Probability of Type 4.3

P(AB;AC;AD;BE) likewise does not factor into already established probabilities. Thus the power curve regression method was employed. The resulting formula was $.23803\gamma^{2.38356}$ with $R^2=.9995$.

Probability of Type 4.4

P(AB;AC;AD;AE) was also attacked with the power curve regression approach. The method worked particularly well with this probability type: $R^2=.9999$. The probability was estimated by $.29211\gamma^{1.85237}$.

Probability of Type 4.5

P(AB;AC;BC;DE) equals P(DE|AB;AC;BC)P(AB;AC;BC). As in Type 4.1, the second factor has already been regarded as essentially zero. In this case, however, we know the magnitude of the overall error due to regarding this as zero. The overall error from Type 7 from S_3 is exactly the same as this type's probability times its type frequency. Thus, we know the overall error due to estimating this quantity as zero is less than .00001.

Probability of Type 4.6

P(AB;BC;CD;DE) again required the power curve regression analysis. The analysis yielded the following estimate: $.16475\gamma^{2.45814}$ with $R^2=.9987$.

Probability of Type 4.7

$P(AB;AC;AD;EF) = P(AB;AC;AD)P(EF) = P(AB;AC;AD)\gamma$ due to independence. Therefore, this probability equals $.34067\gamma^{2.63484}$ since $P(AB;AC;AD)$ was estimated earlier for Type 4 as $.34067\gamma^{1.63484}$.

Probability of Type 4.8

$P(AB;AC;DE;DF) = P(AB;AC)P(DE;DF)$ due to independence. This quantity equals $[P(AB:AC)]^2$ by symmetry. Thus, we can use the estimate: $[\gamma-\alpha_3/3]^2$.

Probability of Type 4.9

$P(AB;AC;BD;EF) = P(AB;AC;BD)P(EF) = P(AB;AC;BD)\gamma$ by independence. This was Type 5 from the previous work, so we can simply multiply the old estimate by γ to obtain: $.27732\gamma^3 + 1.4235\gamma^4$.

Probability of Type 4.10

$P(AB;AC;DE;FG) = P(AB;AC)P(DE)P(FG) = P(AB;AC)\gamma^2$ due to independence. Using the previously established formula for $P(AB;AC)$, we obtain the following expression for this probability type: $[\gamma-\alpha_3/3]\gamma^2$.

Probability of Type 4.11

$P(AB;CD;EF;GH) = P(AB)P(CD)P(EF)P(GH)$ due to independence. This probability is precisely γ^4.

The following table lists the probability of each type. Due to the approximate nature of most of these probabilities only a few decimal points will be shown; also leading to a simpler presentation.

Type 4.1	P(AB;AC;AD;BC)	$.13\gamma^{2.56}$
Type 4.2	P(AB;AC;BD;CD)	$.14\gamma^{2.05}$
Type 4.3	P(AB;AC;AD;BE)	$.24\gamma^{2.38}$
Type 4.4	P(AB;AC;AD;AE)	$.29\gamma^{1.85}$
Type 4.5	P(AB;AC;BC;DE)	0
Type 4.6	P(AB;BC;CD;DE)	$.17\gamma^{2.46}$
Type 4.7	P(AB;AC;AD;EF)	$.34\gamma^{2.63}$
Type 4.8	P(AB;AC;DE;DF)	$[\gamma - \alpha_3/3]^2$
Type 4.9	P(AB;AC;BD;EF)	$.28\gamma^3 + 1.42\gamma^4$
Type 4.10	P(AB;AC;DE;FG)	$[\gamma - \alpha_3/3]\gamma^2$
Type 4.11	P(AB;CD;EF;GH)	γ^4

We must concede that the estimate for S_4 is cruder than for the previous S terms. However, it is not likely to be grossly inaccurate and should provide us with a reasonable upper bound on the error due to omitting it and the subsequent terms. Estimates of S_4 for four through ten samples could be obtained for each γ, α_3 pair (from the Birnbaum-Hall tables) by summing the products of estimates and frequencies. As examples, three γ, α_3 pairs were selected for each value of k samples such that for one α_k was approximately .10, for the second, α_k was approximately .05, and for the third it was approximately .01. The next table presents S_4 and the estimated p-value after each estimated S term is included for each example.

k	S_4	S_1	S_1-S_2	$S_1-S_2+S_3$	$S_1-S_2+S_3-S_4$
4	.000242	.1285	.0972	.1014	.1011
4	.0000406	.0571	.0472	.0482	.0482
4	.00000119	.0110	.0100	.0101	.0101
5	.00162	.1319	.0914	.1007	.0991
5	.000328	.0629	.0488	.0513	.0509
5	.00000899	.0112	.0099	.0101	.0100
6	.00409	.1384	.0886	.1040	.0999
6	.000770	.0638	.0474	.0512	.0504
6	.0000212	.0111	.0097	.0099	.0099
7	.00619	.1322	.0805	.0991	.0929
7	.00128	.0633	.0454	.0504	.0492
7	.0000373	.0112	.0096	.0098	.0098
8	.00968	.1372	.0790	.1034	.0937
8	.00203	.0660	.0455	.0521	.0501
8	.0000638	.0120	.0101	.0104	.0104
9	.0106	.1257	.0711	.0955	.0850
9	.00272	.0661	.0445	.0523	.0496
9	.0000782	.0114	.0096	.0099	.0099
10	.0142	.1294	.0700	.0992	.0850
10	.00312	.0631	.0416	.0500	.0469
10	.0000936	.0110	.0091	.0096	.0095

The last two columns in the preceding table: $S_1 - S_2 + S_3$ and $S_1 - S_2 + S_3 - S_4$ provide an upper and lower bound for the exact p-value respectively. Consequently, if these quantities are very close in magnitude, we have confidence that $S_1 - S_2 + S_3$ which is conservative by design is also an accurate p-value. It appear that $S_1 - S_2 + S_3$ is an excellent estimate of the p-value when it is near .01. Further, it is an excellent estimate when α is near .05 for seven or less samples. For α near .10, $S_1 - S_2 + S_3$ remains an excellent estimate for six or less samples. It is evident, however, that $S_1 - S_2 + S_3$ may be quite conservative for seven or more samples at α=.10, yet it is much more accurate than the Bonferroni S_1.

The above table also illustrates that $S_1 - S_2 + S_3$ is very much superior to either S_1 or $S_1 - S_2$. In particular, we observe that unless α is very small, the Bonferroni estimate is highly conservative.

AN ESTIMATE OF S_4

The following table presents a curve formula for an estimate of S_4 for each value of k, when α_k is less than or equal to .10. These were calculated by means of power curve regression, such that the estimate of S_4 is predicted by γ.

k samples	estimate of S_4
4	$.7664\gamma^{2.1198}$

5	$9.9229\gamma^{2.0385}$
6	$48.8156\gamma^{2.0259}$
7	$151.6750\gamma^{2.0146}$
8	$393.8129\gamma^{2.0127}$
9	$840.1617\gamma^{2.0067}$
10	$1683.2251\gamma^{2.0071}$

CONCLUSION

This examination of the size of the S_4 term again adds support to the claim of accuracy of the curve-estimated p-values. At the same time, it underscores the need for caution in interpreting the curve p-values when the procedure-wise p-value is near .10 for the larger number of samples. Users must be warned about the degree of conservatism under these circumstances.

C H A P T E R   F O U R

ASSESSMENT OF OVERALL CURVE ACCURACY

The process of determining the k-sample p-value by use of the curves presented in Chapter Two has two sources of inaccuracy: 1) the omitted S terms and 2) the approximations used in the probability estimates. The first source was examined in Chapter Three and the second in Chapter Two. Although evaluation of the inaccuracy produced by approximation of some of the unknown probability types is not easily done, there are several ways by which the total inaccuracy of the curves can be assessed. We can compare the p-values yielded by this method with exact p-values for $k=4$, $n<11$ computed by Becker and Taylor (1982), and for other k and n combinations newly computed here. Secondly, we can compare our p-values with those simulated by Gardner et al (1980). Third, we can perform our own simulations.

BECKER AND TAYLOR'S RESULTS (k=4; n:small)

Using Hodges' (1957) computational scheme, Becker and Taylor calculated the exact p-values for k=4 for sample sizes 10 and under. The following table compares their exact p-values with the new curve-estimated p-values and the simple Bonferroni p-values. These estimated p-values were actually calculated from the set of power curve equations. This will also be the case for subsequent analyses.

n	statistic	exact	estimate	Bonferroni
5	1.000	.0402	.0407	.0476
6	1.000	.0118	.0118	.0130
7	.875	.0416	.0417	.0490
7	1.000	.0033	.0033	.0035
8	.750	.0891	.0894	.1119
8	.875	.0135	.0135	.0149
8	1.000	.0009	.0009	.0009
9	.666	.0327	.0327	.0378
9	.777	.0043	.0042	.0044
9	.888	.0002	.0002	.0003
10	.700	.0612	.0613	.0740
10	.800	.0113	.0113	.0123
10	.900	.0013	.0013	.0013

The preceding table indicates that the curve-estimated p-values for these examples are extremely close to the exact values. In addition, the p-values appear to have fairly good accuracy even when they exceed .10. For example, Becker and Taylor report that the exact p-value for n=10; t=.600 is .2200, while we calculate that the curve-estimated p-value is .2215.

MORE COMPARISONS WITH EXACT P-VALUES (k=3-6; n:moderate)

Using the computer algorithm approach, we were able to obtain exact p-values for various n and k combinations up through k equals six. The size of the largest n computed decreases as k increases due to computer memory limitations. The following table presents the curve estimated p-values and also the simple Bonferroni p-value for comparison purposes.

k	n	Statistic	exact	estimate	Bonferroni
3	50	.280	.1000	.1002	.1176
3	50	.300	.0575	.0575	.0651
3	50	.320	.0314	.0314	.0345
3	50	.340	.0163	.0163	.0175
3	100	.200	.0933	.0935	.1092
3	100	.210	.0633	.0634	.0722
3	100	.220	.0419	.0419	.0467
3	100	.230	.0271	.0271	.0296
3	100	.240	.0171	.0171	.0184
3	100	.250	.0105	.0105	.0112
3	100	.260	.0063	.0063	.0067
4	30	.400	.0761	.0762	.0939
4	30	.433	.0340	.0340	.0393
4	30	.466	.0137	.0137	.0152
4	30	.500	.0051	.0051	.0054
4	46	.326	.0711	.0711	.0871
4	46	.348	.0374	.0374	.0435
4	46	.370	.0185	.0185	.0207
4	46	.391	.0086	.0086	.0094
4	46	.413	.0038	.0038	.0040
5	10	.700	.0935	.0944	.1234
5	10	.800	.0180	.0180	.0206
5	16	.563	.0859	.0865	.1120
5	16	.625	.0258	.0259	.0302
6	7	.857	.0898	.0921	.1224
6	7	1.000	.0078	.0079	.0079
6	8	.875	.0308	.0312	.0373
6	8	1.000	.0022	.0022	.0023

The above comparisons exhibit the excellent accuracy of the curve estimated p-values, for a wide variety of examples. We do observe, however, that this accuracy deteriorates slightly as k grows larger for p-values close to .10, as expected. Although not reported in the above table, we again observed that the curve p-values were reasonably close to the exact, when p-values are significantly larger than .10. For example, in the k=6, n=8 case, the exact p-value for statistic .75 is .5535, while the estimated p-value is .5770.

GARDNER ET AL.'S RESULTS

Gardner et al. (1980) published percentiles for T statistics based on simulations of 5000 replications. These can be compared with hypothesized percentiles from the curve-estimated method. The next table compares some of these published percentiles from S (Simulated) with those from C (Curve-Estimated):

		k=4		k=6		k=8		k=10	
n	%	S	C	S	C	S	C	S	C
10	.90	.60	.60	.70	.70	.70	.70	.70	.70
10	.95	.70	.70	.70	.70	.70	.70	.80	.80
10	.98	.70	.70	.80	.80	.80	.80	.80	.80
10	.99	.80	.80	.80	.80	.80	.80	.80	.80
50	.90	.30	.30	.32	.32	.34	.34	.34	.34
50	.95	.32	.32	.34	.34	.36	.36	.36	.36
50	.98	.34	.34	.36	.36	.38	.38	.40	.40
50	.99	.36	.36	.38	.38	.40	.40	.42	.40

100	.90	.21	.21	.23	.23	.24	.24	.25	.25
100	.95	.22	.22	.25	.24	.25	.25	.26	.26
100	.98	.24	.24	.26	.26	.27	.27	.28	.28
100	.99	.25	.26	.28	.27	.28	.29	.29	.29

There is good agreement between these simulated percentiles and those determined from the curves. Out of 48 comparisons, there are four instances above in which the estimates differ by 1/n. Three of these are for the .99 percentile. Gardner et al concede that the confidence interval for the percentiles often straddles a step in the cdf, especially for the .99 percentile. Recall that the curve estimated p-values are probably very accurate around the .01 level. In view of all this, the correspondence between the curve-estimated and simulated percentiles may be as good as the correspondence between the exact and simulated percentiles.

NEW SIMULATIONS

A simulation study with 200,000 replications was conducted. In every replication each of 10 samples had 30 observations. The maximum Kolmogorov-Smirnov statistic for samples A,B,C,D, and E was calculated for each replication to simulate the p-values for k=5, likewise the maximum statistic for samples A,B,C,D,E, and F was used to simulate the p-value for k=6, and so forth. The following table presents exact p-values (E) when available, the simulated p-values (S), the lower (L) and (U) upper bounds for a 95%

confidence interval for each simulated p-value (with no adjustment for multiple comparisons), the curve estimated p-values (C), and the hypothesized p-value using only terms S_1, known generally as the Bonferroni approximation (B).

T	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
.433								
E	.0182	.0340						
S	.0185	.0347	.0537	.0751				
L	.0179	.0339	.0527	.0740				
U	.0191	.0355	.0546	.0763				
C	.0182	.0340	.0532	.0757				
B	.0196	.0393	.0655	.0982				
.467								
E	.0072	.0137						
S	.0074	.0143	.0226	.0322	.0430	.0547	.0675	.0810
L	.0070	.0137	.0220	.0315	.0421	.0537	.0664	.0798
U	.0078	.0148	.0233	.0330	.0439	.0557	.0686	.0822
C	.0072	.0137	.0219	.0317	.0429	.0557	.0700	.0864
B	.0076	.0152	.0253	.0380	.0531	.0708	.0911	.1139
.500								
E	.0026	.0051						
S	.0026	.0053	.0085	.0124	.0167	.0216	.0269	.0324
L	.0024	.0049	.0081	.0119	.0161	.0209	.0262	.0316
U	.0028	.0056	.0089	.0129	.0173	.0222	.0276	.0331
C	.0026	.0051	.0082	.0119	.0163	.0213	.0267	.0331
B	.0027	.0054	.0090	.0135	.0189	.0252	.0324	.0405
.533								
E	.0009	.0017						
S	.0009	.0018	.0029	.0043	.0060	.0077	.0098	.0120
L	.0008	.0016	.0027	.0040	.0057	.0073	.0094	.0115
U	.0010	.0020	.0031	.0046	.0063	.0081	.0103	.0124
C	.0009	.0017	.0028	.0041	.0056	.0073	.0093	.0114
B	.0009	.0018	.0029	.0044	.0062	.0082	.0106	.0132
.567								
S			.0009	.0013	.0019	.0024	.0030	.0038
L			.0008	.0012	.0017	.0022	.0028	.0035
U			.0010	.0015	.0021	.0026	.0033	.0040
C			.0008	.0012	.0017	.0023	.0029	.0035
B			.0009	.0013	.0018	.0024	.0031	.0039

The match between the simulated and curve-estimated p-values is very good for most comparisons. There are several places where the curve estimate is just slightly below the confidence interval. But comparisons between the exact p-values, where available, and the simulated p-values indicate that the simulated p-values are slightly high. Not surprisingly, the curve estimated p-values which are clearly outside the confidence intervals are those for k of nine or more with p-values greater than .05. We recognize that we are performing many simultaneous tests, so we would expect about 5% of the confidence intervals would not include the true p-value. However, since we expected an inaccuracy a priori in these cases, it seems likely that these instances reflect real differences between the true and the estimated p-values. The degree of conservatism shown for these cases is not surprising in view of what we observed in the previous chapter about the fourth S-term. Thus, this simulation study tends to confirm the overall accuracy of the curve approach, with the previously established exceptions.

A second set of 200,000 simulations was then performed, however, this time the sample size for each sample was 10 observations. The following table compares p-values in the same manner used for n=30. Examination of the table reveals the same pattern as observed for the n=30 case. All curve-estimated p-values were within the corresponding confidence intervals (or slightly outside in one instance), with the clear exception of the p-values greater than .05 for nine and ten samples.

T	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
.800								
E	.0059	.0113	.0180					
S	.0059	.0110	.0180	.0259	.0348	.0448	.0552	.0668
L	.0056	.0106	.0174	.0252	.0340	.0439	.0542	.0657
U	.0062	.0115	.0186	.0266	.0356	.0457	.0562	.0668
C	.0059	.0113	.0180	.0260	.0354	.0460	.0579	.0714
B	.0062	.0123	.0206	.0309	.0432	.0576	.0740	.0926
.900								
E	.0007	.0013	.0020					
S	.0006	.0013	.0021	.0031	.0041	.0054	.0069	.0085
L	.0005	.0011	.0019	.0028	.0038	.0050	.0065	.0085
U	.0007	.0014	.0023	.0033	.0044	.0057	.0073	.0089
C	.0006	.0013	.0020	.0030	.0042	.0055	.0069	.0085
B	.0007	.0013	.0022	.0033	.0046	.0061	.0078	.0098

Both simulation studies provide additional reassurance that the curve-estimated p-values are very close to the true p-values. They also alert us again to the conservatism of the nine and ten sample p-values.

CHAPTER SUMMARY

The exact p-values for four samples published by Becker and Taylor are extremely close to the curve estimated p-values. The exact p-values computed here for up to six samples are also remarkably close to the curve-estimated p-values.

The percentiles simulated by Gardner et al. match the curve estimated percentiles closely; the few discrepancies were of size 1/n. Moreover, this serves as the first check of the accuracy of the curves for a large sample size (in this case, n=100).

Our simulation studies revealed no problems with the accuracy of the curve-estimated p-values, with the possible exception of the nine and ten sample case.

Overall, the curve estimated p-values have passed the above three scrutinies.

CHAPTER FIVE

EXTENSION OF THE PROCEDURE TO UNEQUAL SAMPLE SIZES

At the beginning of Chapter Two, it was mentioned that

$$T = \sup_{x,i,h} |F_i(x) - F_h(x)|$$

was not very appropriate for unequal sample sizes. The reason for this is best illustrated with an example. Say there are three samples denoted A, B and C with respective sample sizes 100, 100 and 5. The two-sample statistic comparing samples A and B is based on two large sample sizes while that comparing samples A and C is based on one large and one small sample. Thus the AB sample comparison is a more precise measure of its true difference than the other two comparisons. The statistic T, however, treats all three statistics identically. It does not take into account which comparison has resulted in the maximum difference. With this scheme, the variability of the distribution of T introduced by the AC and BC comparisons will reduce the statistic's sensitivity to the difference between A and B. A statistic which weights each comparison by the sample sizes of its pairs would be much more appropriate. We propose the following statistic, U, as an unequal sample size analogue of T:

$$U = \max_{i,h} [n_i n_h/(n_i+n_h)]^{1/2} \sup_x |F_i(x) - F_h(x)|$$

Asymptotically, each comparison of sample pairs has the same null distribution when the above sample size weights are applied. Note: U reduces to a constant times T in the equal sample size setting, and consequently has equivalent distribution.

DISTRIBUTION OF U

For sake of simplicity, let us first consider the k=3 samples case. Again, let the samples be labelled A, B and C; now AB represents the weighted comparison between A and B.

$$\begin{aligned}
&P(U \geq u) = \\
&P(AB \geq u \text{ or } AC \geq u \text{ or } BC \geq u) = \\
&P(AB \geq u) + P(AC \geq u) + P(BC \geq u) \\
&-[P(AB;AC \geq u) + P(AB;BC \geq u) + P(AC;BC \geq u)] \\
&+ P(AB;AC;BC \geq u)
\end{aligned}$$

As in the equal sample size case, the first three terms are simply all of the possible two-sample p-values. Now let us reconsider the unequal sample sizes 100, 100, and 5.

Now, unfortunately the remaining terms are not sums of equal probability types. In other words, even though $P(AB \geq u)$, $P(AC \geq u)$, and $P(BC \geq u)$ are asymptotically equal, the type 2 joint probabilities such as $P(AB;AC \geq u)$ and $P(AC;BC \geq u)$ are not necessarily equal. This is related to the fact that in the first probability, the compared pairs share a large sample and in the second they share a small sample. Since the type 2 probabilities are not equal, it is unlikely that a simple scheme such as that developed for equal sample sizes is possible.

As a consequence, we must turn to a somewhat different approach. By summing the two sample p-values for a specified critical value U for every $(C_{k,2})$ sample size pair we can estimate S_1 exactly. In the equal sample size setting we plugged γ into a function of the form $a\gamma^b$ to take S_2 and S_3 into account. Let us propose that the mean of the $(C_{k,2})$ p-values obtained can be labelled γ and then plugged into the same function to adjust for S_2 and S_3. We now need to understand, as well as possible, what this does and to see how effectively this approximates the exact S_2 and S_3 terms.

Therefore, in the unequal sample size case, we will consider the use of the following approach. We could average all of the $(C_{k,2})$ two-sample p-values for a specified critical value U using every sample size pair, and then insert this quantity into our k-sample formula. This scheme has the following result: S_1 is exactly estimated and S_2 and S_3 are approximately estimated by a function of the mean of the two-sample p-values for U. It is important to realize that this strategy has the desirable property that it reduces to the equal sample size method if the sample sizes are in fact equal. Furthermore, once the sample sizes are sufficiently large so that the asymptotic distribution formula applies, all the two-sample p-values would be equal, so that the corresponding k-sample p-value will be the same as it would be for equal sample sizes corresponding to the same critical value.

Before we investigate the validity of this idea in general, we shall make some of these concepts more concrete by concentrating on a single example. In this example, there are four samples A, B, C, and D with sizes 5, 10, 15, and 20 respectively. Using the computer path algorithm in conjunction with elementary probability laws we can obtain exact relevant probabilities. Using a critical value of 1.50, we observe the following under the null hypothesis:

Type 1	P(AB)=.003996
	P(AC)=.008772
	P(AD)=.012309
	P(BC)=.010033
	P(BD)=.012447
	P(CD)=.013635
	Mean=.010199

Type 2	P(AB;AC)=.000956
	P(AB;AD)=.001252
	P(AB;BC)=.000342
	P(AB;BD)=.000421
	P(AC;AD)=.002586
	P(AC;BC)=.000324
	P(AC;CD)=.000539
	P(AD;BD)=.000400
	P(AD;CD)=.000495
	P(BC;BD)=.001686
	P(BC;CD)=.000969
	P(BD;CD)=.000731
	Mean=.000892
Type 3	P(AB;CD)=.000054
	P(AC;BD)=.000109
	P(AD;BC)=.000123
	Mean=.000096
Type 4	P(AB;AC;AD)=.000559
	P(AB;BC:BD)=.000121
	P(AC;BC;CD)=.000071
	P(AD;BD;CD)=.000050
	Mean=.000200
Type 5	P(AB;AC;BD)=.000032
	P(AB;AC;BC)=.000036

P(AD;AC;BC)=.000044

P(AB;AD;CD)=.000023

P(AC;AD;BD)=.000044

P(AB;BC;CD)=.000014

P(BD;BC;AC)=.000028

P(BD;BC;AD)=.000032

P(BD;AB;CD)=.000012

P(AC;CD;BD)=.000017

P(BC;CD;AD)=.000022

P(AB;AC;CD)=.000022

Mean=.000027

Type 6 P(AB;AC;BC)=.000008

P(BC;CD;BD)=.000012

P(AD;CD;AC)=.000023

P(AB;BD;AD)=.000016

Mean=.000014

DISCUSSION OF EXAMPLE'S PROBABILITY TYPES

In the equal sample size setting, recall that within each type all probabilities are equal, but clearly this is not the case here. We see in Type 1 that the single probabilities vary from .004 to .012 for these small sample sizes. This illustrates that the asymptotic distribution does not really apply to small samples. For large samples, however, these probabilities should be very close to equal.

For Type 2, we observe that when both pairs share a sample with a relatively small sample size, they are more dependent than when they share a sample with a relatively large size ("relative" means in contrast to other samples in the comparison). For example, P(AD)=.012, P(BD)=.012, and P(AB)=.004, but, P(BD|AD)=.03 and P(AB|AD)=.10. Thus, the magnitude of the Type 2 probability depends strongly on the relative size of the shared sample. Those Type 2 probabilities which share the smallest sample A are approximately twice as large as the probabilities which involve the sharing of the largest sample D.

We can see precisely the same phenomenon in Type 4. Those probabilities which involve terms which have the smallest sample in common are larger than those which have terms with the largest sample in common.

For Types 5 and 6, the probabilities are a little more stable, presumably since the components share several samples. For these types it is difficult to observe any patterns such as those described for earlier types.

In conclusion, we observe that within each probability type, there is variability due in part to the relative size of those samples which appear in more than one pair. In other words, the degree of dependency of the pairs is determined by this relative size.

For this same example, let us consider how well a modified equal sample size strategy can work for the unequal sample size situation. We average all the two-sample p-values associated with the critical value U to get a single γ, which represents the typical p-value. If this averaging process works, when we plug this mean γ into each probability type formula, we should get a value similar to the average of the exact probabilities for that type. If they are in fact equal then this probability type is estimated without error.

For this example, γ=.010199. Our Type 2 formula is $\gamma-\alpha/3$. However, in this setting, we do not have the individual α probabilities. We do have from Chapter Two, an estimate for α as a function of γ. Using this we obtain the following estimate: P(Type 2)=$.5245\gamma^{1.3916}$. Plugging γ into this formula, the answer is .000888. This closely approximates the exact average probability which was .000892. Thus, the exact probabilities are especially high when the pairs share small samples and especially low when they share large samples, but they average out to that probability they would have had if all sample sizes were equal.

Inserting γ into the Type 3 formula of γ^2, we obtain .000104. This is close to the exact average of .000096. Note, this approximation will match the exact if all of the two-sample p-values are equal.

Using the Type 4 formula, the approximate probability is .000189. The exact average probability equals .000200. Again all the specific Type 4 probabilities appear to average out to about the same probability we would expect with equal sample sizes.

Our estimate for Type 5 is .000030. The exact average for this probability if .000027. Therefore, again the approximation does a reasonable job.

The following table summarizes the results taking into account the frequency of each type for this example.

Type	Freq	Estimate	Exact	Estimate - Exact
1	6	.061192	.061192	.000000
[-2 + 6]*	12;4	-.010656	-.010644	-.000012
-3	3	-.000312	-.000287	-.000025
+4	4	.000756	.000802	-.000054
+5	12	.000364	.000327	.000037
+7	0	-	-	-
+8	0	-	-	-

* Recall that the Type 2 estimate is actually a joint estimate of both Types 2 and 6.

Thus, we see, for this example, that each individual probability type is well approximated by this averaging approach. Of course, we do not need to estimate each type; we can simply plug ¥ into the four sample curve to obtain an estimate.

Let us redo the problem as the user would do it. Say we have the same four samples, and that the U statistic obtained as the maximum of the weighted statistics of the six sample pairs was 1.50. Using a table or brief computer routine, we then determine the following p-values for 1.5 for each sample size combination. We obtain: .012309, .012447, .013635, .003996, .008772, and .010033. We then take the mean of these to get .010199. This is inserted into the regular k=4 samples curve: $6\gamma - 5.3761\gamma^{1.3755}$ which yields .05139. This is our final procedure-wise p-value. (Note: using the computer algorithm we calculated that the exact p-value is .05134.)

PROPOSED UNEQUAL SAMPLE SIZE PROCEDURE

There is no guarantee that this method will work as well for other sample size combinations. However, it certainly suggests that this may be a reasonable approach. Therefore, in general, we can try using the following procedure to determine a k-sample p-value estimate. We calculate the U statistic. Then, using a table, brief computer routine, or the asymptotic formula if the sample sizes are sufficiently large, we obtain p-values for all $C_{k,2}$ pairs of sample sizes for that critical value U. (Note: when using the asymptotic formula, for each pair of sample sizes, use the smallest possible value of the two-sample statistic which is greater than or equal to the k-sample statistic U. This will produce a much more accurate, i.e., less conservative, answer than plugging the k-sample statistic U into the asymptotic formula when it is not a possible two-sample statistic. When using the path p-value algorithm, either value of U will produce that correct possible p-value.) We then average these p-values and insert this average into the k-sample curve which appears in Chapter Two, to yield the final answer.

As an alternative, for a slightly more precise answer, the user could actually work out each probability type separately, using the probability types from Chapter Two as done for the example in this chapter. With this approach, he or she could incorporate the exact types for Types 3 and 8. This, however, is not likely to make more than a minimal difference.

ACCURACY OF ESTIMATED P-VALUES FOR UNEQUAL SAMPLE SIZES

Clearly, we need to determine how well this strategy works for a wide variety of sample size combinations. A good way to evaluate the accuracy of the proposed method is to compare the resulting estimates with the corresponding exact p-values for specific sample sizes. Unfortunately, we are restricted to a small number of samples and relatively small sample size, due to computer memory limitations. The following pages compare exact and approximated p-values for examples of three to six samples. If the method works for six samples, it is likely it works for higher numbers of samples. For the six-sample p-values estimates to well approximate the exact values, the averaging approach for all types has to be effective. However, it may be that the total error gradually increases with increasing number of samples, just as it does in the equal sample size case.

The following table provides a quick summary of the exact and curve approximated p-values, and the Bonferroni estimates at about the .10, .05 and .01 significance levels.

			U	Exact	Curve	Bonferroni
5	10	15	1.2909	.1050	.1030	.1210
			1.4200	.0530	.0539	.0608
			1.6329	.0114	.0115	.0122
10	15	30	1.3063	.1070	.1088	.1284
			1.4605	.0536	.0541	.0611
			1.6865	.0102	.0102	.0109
8	17	53	1.3219	.1004	.1036	.1218
			1.4425	.0497	.0512	.0576
			1.6602	.0101	.0102	.0109
15	15	45	1.3416	.1016	.1013	.1190
			1.4605	.0553	.0565	.0640
			1.7143	.0092	.0092	.0098
15	45	45	1.3703	.0904	.0911	.1062
			1.4757	.0547	.0555	.0628
			1.7143	.0099	.0103	.0109
30	40	50	1.3567	.0981	.0985	.1154
			1.4836	.0492	.0494	.0555
			1.7206	.0104	.0105	.0111
20	40	80	1.3555	.1025	.1044	.1228
			1.4846	.0477	.0482	.0541
			1.7344	.0101	.0103	.0110
75	75	150	1.3880	.0943	.0955	.1116
			1.5085	.0493	.0495	.0556
			1.7442	.0106	.0107	.0113
50	100	150	1.3685	.1023	.1041	.1225
			1.5011	.0492	.0503	.0566
			1.7555	.0095	.0096	.0102

				U	Exact	Curve	Bonferroni
10	10	15	15	1.4605	.0893	.0888	.1111
				1.5513	.0502	.0506	.0601
				1.7888	.0090	.0090	.0098

8	10	12	14	1.4013	.0981	.0981	.1240
				1.5130	.0492	.0489	.0581
				1.7249	.0099	.0099	.0108
5	10	15	20	1.4150	.1004	.1011	.1283
				1.5000	.0513	.0514	.0612
				1.7000	.0105	.0104	.0114
10	10	10	30	1.4605	.0811	.0813	.1007
				1.5652	.0454	.0464	.0548
				1.7885	.0092	.0094	.0102
10	10	30	30	1.4200	.1029	.1037	.1319
				1.5518	.0519	.0543	.0650
				1.7885	.0090	.0092	.0100
15	15	15	50	1.4605	.1021	.1053	.1341
				1.5399	.0486	.0492	.0583
				1.7890	.0099	.0101	.0111
9	19	29	39	1.4258	.1008	.1032	.1312
				1.5387	.0497	.0509	.0605
				1.7672	.0100	.0100	.0112
20	20	50	60	1.4362	.0994	.1011	.1282
				1.5667	.0492	.0506	.0602
				1.7930	.0100	.0103	.0112
20	40	60	80	1.4605	.1002	.1032	.1312
				1.5614	.0480	.0491	.0583
				1.8074	.0103	.0106	.0116

					U	Exact	Curve	Bonferroni
10	10	10	20	20	1.5492	.0926	.0948	.1240
					1.5811	.0398	.0394	.0472
					1.8074	.0107	.0106	.0118
10	10	15	15	20	1.4697	.1074	.1086	.1447
					1.5652	.0530	.0533	.0657
					1.8054	.0100	.0099	.0110
10	10	15	20	30	1.4757	.0992	.1006	.1326
					1.5811	.0489	.0492	.0602
					1.8054	.0100	.0099	.0110

						U	Exact	Curve	Bonferroni
5	5	5	10	10	15	1.4697	.0924	.0923	.1227
						1.5811	.0529	.0550	.0690
						1.7889	.0103	.0101	.0113

5	5	5	15	15	15	1.5492	.0887	.0943	.1257
						1.5811	.0554	.0565	.0712
						1.8074	.0112	.0111	.0125
5	5	10	10	10	20	1.5000	.0971	.0977	.1309
						1.5811	.0484	.0507	.0632
						1.8000	.0102	.0100	.0112

CONCLUSIONS ON THE METHOD

For the sample size combinations listed above, the estimated p-values differ from the exact p-values by at most .005 at the .10 level, .003 at the .05 level, and .0004 at the .01 level. Usually the estimates exceeded the exact p-values. In those cases where the estimates underestimated the p-value, the error was small. These estimates are not as accurate as estimates for equal sample sizes, however, they are not very far off and are clearly much superior to a simple Bonferroni estimate.

It is interesting to observe that for two of the three cases when k=6, the conservatism is less than we would expect with equal sample sizes. This observation, along with further examination of the above tables, suggests the possibility that the conservatism does not grow with k as much in the unequal sample size setting as it does with equal sample sizes. In fact, the degree of the conservatism may stay relatively constant across k, when the samples are unequal.

There are not any striking patterns in the above tables. The method works well for a wide variety of sample size combinations. However, the degree of the accuracy varies somewhat. The worst case above for the .05 level is the 10,10,30,30 combination. The worst case for the .10 level is the 5,5,5,15,15,15 combination.

The above tables add credibility to the proposed unequal sample size approach. We can say that it appears to be working quite well for three to six samples. Furthermore, the fact that it works for six samples suggests that the averaging method for each probability type is effective, and probably should continue to be effective for a larger number of samples.

REMARKS ON ASCERTAINMENT OF TWO SAMPLE P-VALUES

This method clearly relies on the availability of two-sample p-values. However, this availability is not as good as we would like. Kim and Jennrich (1970) tabulate all p-values for all possible combinations of unequal sample sizes up to 25. Birnbaum and Hall (1960) have two-sample p-values for all equal sample sizes up through 40. When these tables are unavailable or are exceeded, the user has two choices: write a brief computer program to get the exact p-value or resort to the asymptotic formula. The computer program is simple and inexpensive for the two-sample case, but obviously this task will not appeal to some, if not most, users. The alternative is the use of the asymptotic formula to get an approximate p-value. Again, when employing the asymptotic formula, the user must be reminded to use the smallest possible value of U for a specific two sample size combination which is greater than or equal to the k-sample statistic.

As far as can be determined, there has not been a great deal of attention paid to the degree of accuracy of the asymptotic formula. We have investigated this question somewhat, though certainly not exhaustively. We observe that the asymptotic formula is not bad at all for equal sample sizes greater than forty. Thus, in conjunction with the Birnbaum-Hall tables, the task of determining p-values for the equal sample size case is essentially complete. Unequal sample sizes pose more difficulty, however. Even when both samples exceed 100, the unequal sample size p-value is not very precisely approximated by the asymptotic distribution. For example, when the sample sizes are 130 and 120, the statistic U=1.46847 has the exact p-value of about .023, whereas the asymptotic formula indicates that the p-value is .027. Therefore, the unequal sample size case requires much more tabulation or an improved modification of the asymptotic formula.

This conclusion contradicts that of Kim (1969) who suggested that the error from the asymptotic formula is worse when n=km, where n is one sample size, m is the other, and k is any integer, than when n does not equal km for any k. He determined those exact p-values which corresponded to the smallest values of U greater than or equal to 1.225, 1.358, 1.628, which corresponded to 10%, 5% and 1% levels of the U distribution. He then compared these p-values to .10,.05,.01 and then drew the conclusion about integer multiples. However, he should have compared these p-values with the p-values of the exactly corresponding U. For example, when n=100 and m=30, the p-value corresponding to U of 1.36109 is .040, and when n=100 and m=100, the p-value corresponding to U of 1.41421 is .036. Since these U values are the smallest values greater than or equal to 1.358 for their respective sample sizes, Kim compared these p-values to .05 and found the 100,30 combination superior to the equal sample size case. However, the p-value corresponding to U=1.36109 is about .049, whereas the p-value corresponding to U=1.41421 is about .037. So, in fact, the p-value in the 100, 100 case, .036, was much closer to its asymptotic p-value, .037, than the p-value in the the 100,30 case, .040, was to its asymptotic p-value, .049.

Kim and Jennrich (1970) suggest using the following modification for the asymptotic formula: add one half divided by the square root of n to U, where n is the larger of the two sample sizes. Let us refer to this method as MOD1. Another modification which can be developed as an obvious generalization of Harter's (1980) one sample asymptotic formula (MOD2, say), appears to work very well in some situations. In U, we replace n by $n + ((n+4)^{.5})/3.5$ and likewise we replace m by the analogous correction. We investigated the p-values (all under .025 since this is the area of concern for this method) obtained by the asymptotic formula and the two modifications, for 18 different sample size combinations. The conclusion was that for equal sample sizes, the asymptotic formula was the clear winner. However, for unequal sample sizes, MOD2 always more closely approximated the exact p-value than the asymptotic formula, and in some cases, was extremely close to exact. MOD1 appeared to work better than MOD2 when the two samples were relatively prime. However, these conclusions are very tentative; they need to be more thoroughly investigated. The following examples for two different sample size combinations illustrate these tentative conclusions:

				Estimated P-values		
n	m	U	exact	Asymptotic	MOD1	MOD2
100	40	1.469	.0223	.0269	.0199	.0224
100	40	1.496	.0190	.0229	.0169	.0189
100	40	1.523	.0160	.0195	.0143	.1060
53	17	1.441	.0223	.0318	.0211	.0239
53	17	1.481	.0165	.0252	.0165	.0186
53	17	1.556	.0100	.0158	.0102	.0113
53	17	1.660	.0051	.0081	.0051	.0056

We noticed that for a given sample size combination, the asymptotic formula can be improved upon greatly by replacing U with a linear transformation of U. This linear transformation can be determined empirically ordinary least squares regression. First we determined the value of U* which would yield the correct p-value using the asymptotic formula, for a series of p-values associated with that sample size combination. Then regression was used to find the linear combination of U which best predicts U*, i.e., U* = a + b U. We calculated the required linear transformations for all sample size combinations up to 50, excluding those combinations whose full distributions are already tabled elsewhere. These regressions were only computed for small p-values $p<.03$ and may not apply along the entire distribution. The fits were very good; in more than two-thirds of the regressions, $R^2>.999$. This tabulation appears in Appendix 5.1. The first entry represents alpha and the second entry, beta. For example, for sample sizes 24 and 28, a possible value of U is 1.41227, so we must transform U into -.013 + 1.049(U) = 1.468. The exact p-value for this critical value is .0267, when U is placed into the asymptotic formula, the estimated p-value is .0370. However, when 1.468 is plugged into the asymptotic formula, the resulting p-value is .0268. Thus, the table in the appendix is essentially a compact table of p-values.

RECOMMENDATIONS

Recommendations regarding the determination of two-sample p-values are the following. If one plans to make extended use of this k-sample Kolmogorov-Smirnov method, it would be very valuable to write the computer program which calculates the exact two-sample p-values, if this is at all feasible. If not, one should use the available tables for small sizes, including those in the appendix here for the unequal sample size tabulations up to 50. When tables are unavailable or exceeded, one should use the modified asymptotic approaches described above when one has unequal sample sizes. If the sample sizes are equal, one should use the standard asymptotic formula.

If one decides to rely completely on the asymptotic two-sample p-values (and ignore all suggested refinements), a single table which summarizes the U statistic significant at various levels would be helpful. This table applies to both the equal and unequal sample size case. But again, the user must be cautioned that the values in this table are very conservative, if the sample sizes are unequal, unless the sample sizes in each sample exceed about 200. For example, for the moderate sample size combination of 50, 40, 30, the exact p-value for U of 1.48364 is .0492. The curve method using the exact two-sample p-values yields .0494. However, the curve method using the asymptotic two-sample p-values yields .0645. Each entry in the following table is the U statistic associated with the asymptotic p-value labeled on the left and the number of samples labeled on the top.

α	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
.10	1.22	1.40	1.51	1.58	1.64	1.69	1.74	1.77	1.80
.09	1.25	1.42	1.53	1.60	1.66	1.71	1.75	1.79	1.82
.08	1.27	1.44	1.55	1.62	1.68	1.73	1.77	1.81	1.84
.07	1.29	1.47	1.57	1.65	1.70	1.75	1.79	1.83	1.86
.06	1.32	1.50	1.60	1.67	1.73	1.78	1.82	1.85	1.88
.05	1.36	1.53	1.63	1.70	1.76	1.80	1.84	1.88	1.91
.04	1.40	1.57	1.67	1.74	1.79	1.84	1.88	1.91	1.94
.03	1.45	1.61	1.71	1.78	1.83	1.88	1.92	1.95	1.98
.02	1.52	1.68	1.77	1.84	1.89	1.94	1.97	2.00	2.03
.01	1.63	1.78	1.87	1.94	1.99	2.03	2.06	2.09	2.12
.005	1.73	1.88	1.96	2.03	2.07	2.11	2.15	2.18	2.20
.001	1.95	2.08	2.16	2.22	2.26	2.30	2.33	2.36	2.38

APPENDIX 5.1

LINEAR CORRECTION TO ASYMPTOTIC FORMULA

	26		27		28		29	
5	-0.194	1.210	-0.194	1.213	-0.183	1.209	-0.192	1.216
6	-0.168	1.186	-0.242	1.232	-0.142	1.172	-0.114	1.157
7	-0.124	1.152	-0.118	1.151	-0.158	1.154	-0.118	1.144
8	-0.069	1.111	-0.093	1.126	-0.127	1.141	-0.089	1.123
9	-0.066	1.108	-0.124	1.119	-0.063	1.100	-0.064	1.104
10	-0.050	1.092	-0.045	1.088	-0.057	1.096	-0.088	1.117
11	-0.050	1.089	-0.059	1.096	-0.036	1.080	-0.049	1.089
12	-0.020	1.069	-0.059	1.090	-0.059	1.087	-0.047	1.085
13	-0.088	1.080	0.032	1.033	-0.045	1.083	-0.040	1.080
14	-0.056	1.088	-0.039	1.079	-0.068	1.066	-0.052	1.083
15	-0.017	1.062	-0.040	1.074	-0.017	1.062	-0.032	1.071
16	-0.031	1.070	-0.013	1.058	-0.034	1.066	-0.015	1.059
17	-0.016	1.060	-0.013	1.058	-0.017	1.060	-0.019	1.060
18	-0.020	1.060	-0.046	1.062	-0.025	1.063	-0.017	1.058
19	-0.013	1.056	-0.009	1.053	-0.007	1.052	-0.004	1.049
20	-0.021	1.060	-0.013	1.055	-0.022	1.056	-0.010	1.052
21	-0.014	1.055	-0.016	1.054	-0.025	1.051	-0.010	1.052
22	-0.012	1.054	-0.006	1.049	-0.012	1.052	-0.007	1.049
23	-0.009	1.052	-0.007	1.050	-0.004	1.047	-0.004	1.047
24	0.011	1.038	-0.010	1.049	-0.013	1.049	-0.007	1.048
25	-0.070	1.089	-0.001	1.047	0.002	1.043	-0.004	1.046
26	-0.066	1.049	0.025	1.028	-0.021	1.057	0.002	1.042
27			-0.059	1.044	0.012	1.037	0.000	1.043
28					-0.054	1.041	0.024	1.030
29							-0.056	1.042

	30		31		32		33	
5	-0.295	1.268	-0.190	1.210	-0.202	1.219	-0.198	1.219
6	-0.194	1.189	-0.148	1.171	-0.133	1.165	-0.204	1.208
7	-0.111	1.142	-0.130	1.156	-0.125	1.151	-0.132	1.157
8	-0.089	1.124	-0.080	1.121	-0.132	1.132	-0.098	1.126
9	-0.093	1.119	-0.074	1.111	-0.065	1.103	-0.097	1.121
10	-0.107	1.104	-0.013	1.066	-0.067	1.102	-0.058	1.097
11	-0.049	1.088	-0.058	1.094	-0.085	1.112	-0.096	1.095
12	-0.062	1.083	-0.036	1.077	-0.057	1.086	-0.052	1.084
13	-0.029	1.072	-0.032	1.073	-0.030	1.072	-0.027	1.068
14	-0.026	1.068	-0.027	1.068	-0.036	1.073	-0.027	1.068
15	-0.071	1.067	-0.018	1.061	-0.020	1.063	-0.038	1.071
16	-0.032	1.069	-0.007	1.056	-0.065	1.062	-0.038	1.071
17	-0.015	1.057	-0.019	1.060	-0.021	1.061	-0.006	1.051
18	-0.030	1.057	-0.013	1.055	-0.017	1.057	-0.021	1.057
19	-0.010	1.053	-0.012	1.053	-0.010	1.052	-0.012	1.053
20	-0.039	1.055	-0.008	1.050	-0.019	1.053	-0.008	1.050
21	-0.016	1.052	-0.005	1.048	-0.014	1.052	-0.014	1.051
22	-0.009	1.050	-0.004	1.046	-0.010	1.050	-0.030	1.047
23	-0.004	1.046	-0.001	1.044	-0.004	1.046	-0.003	1.045
24	-0.018	1.047	-0.001	1.043	-0.020	1.045	-0.010	1.046
25	-0.013	1.046	-0.007	1.047	-0.003	1.044	-0.008	1.047
26	-0.006	1.046	0.001	1.041	-0.003	1.043	-0.002	1.043
27	-0.011	1.047	-0.002	1.043	0.002	1.040	-0.004	1.042
28	0.008	1.037	-0.002	1.043	-0.006	1.042	0.000	1.041
29	-0.028	1.061	0.000	1.042	0.007	1.037	0.004	1.038
30	-0.051	1.038	0.020	1.028	-0.010	1.047	0.002	1.037
31			-0.053	1.039	-0.007	1.045	0.007	1.037
32					-0.049	1.036	-0.017	1.052
33							-0.050	1.037

	34		35		36		37	
5	-0.211	1.229	-0.286	1.264	-0.219	1.228	-0.213	1.227
6	-0.183	1.198	-0.185	1.201	-0.217	1.205	-0.156	1.177
7	-0.121	1.151	-0.160	1.158	-0.097	1.132	-0.119	1.147
8	-0.083	1.120	-0.111	1.138	-0.121	1.138	-0.086	1.121
9	-0.088	1.119	-0.045	1.094	-0.103	1.109	-0.075	1.107
10	-0.057	1.095	-0.083	1.104	-0.059	1.096	-0.053	1.092
11	-0.032	1.074	-0.050	1.088	-0.057	1.092	-0.045	1.084
12	-0.049	1.085	-0.032	1.076	-0.078	1.081	-0.040	1.077
13	-0.031	1.072	-0.032	1.072	-0.028	1.069	-0.036	1.075
14	-0.037	1.074	-0.045	1.068	-0.032	1.070	-0.032	1.071
15	-0.020	1.062	-0.036	1.065	-0.034	1.068	-0.027	1.066
16	-0.031	1.068	-0.014	1.057	-0.031	1.063	-0.018	1.059
17	-0.060	1.058	-0.025	1.062	-0.008	1.052	-0.022	1.060
18	-0.010	1.052	-0.009	1.053	-0.056	1.055	0.027	1.028
19	-0.013	1.054	-0.011	1.052	-0.021	1.058	-0.007	1.049
20	-0.016	1.054	-0.019	1.051	-0.016	1.050	-0.003	1.046
21	-0.009	1.049	-0.022	1.049	-0.014	1.050	-0.004	1.045
22	-0.010	1.048	-0.008	1.048	-0.011	1.049	-0.002	1.043
23	-0.014	1.051	-0.000	1.042	-0.002	1.043	-0.002	1.043
24	-0.006	1.045	-0.002	1.043	-0.027	1.044	-0.003	1.043
25	0.000	1.041	-0.011	1.043	-0.002	1.042	-0.006	1.044
26	-0.005	1.044	-0.002	1.042	-0.002	1.041	-0.003	1.042
27	0.001	1.040	0.002	1.039	-0.014	1.039	-0.000	1.040
28	-0.003	1.042	-0.010	1.039	-0.005	1.039	0.002	1.038
29	0.000	1.040	0.002	1.038	0.005	1.036	0.003	1.037
30	0.001	1.039	-0.005	1.038	-0.006	1.037	0.004	1.036
31	0.004	1.037	0.003	1.037	0.004	1.036	0.007	1.034
32	0.014	1.031	0.003	1.037	-0.001	1.036	0.007	1.034
33	0.017	1.031	0.004	1.037	0.001	1.036	0.007	1.034
34	-0.047	1.035	0.063	1.001	-0.008	1.044	0.010	1.032
35			-0.043	1.032	0.035	1.016	0.001	1.038
36					-0.040	1.030	-0.033	1.058
37							-0.042	1.031

	38		39		40		41	
5	-0.209	1.227	-0.219	1.235	-0.302	1.276	-0.222	1.231
6	-0.161	1.183	-0.211	1.213	-0.193	1.205	-0.137	1.171
7	-0.128	1.154	-0.128	1.152	-0.139	1.161	-0.134	1.158
8	-0.094	1.127	-0.066	1.111	-0.123	1.129	-0.096	1.125
9	-0.065	1.104	-0.088	1.115	-0.078	1.113	-0.074	1.108
10	-0.074	1.106	-0.054	1.095	-0.093	1.099	-0.056	1.092
11	-0.050	1.088	-0.047	1.085	-0.054	1.090	-0.048	1.086
12	-0.046	1.083	-0.050	1.083	-0.057	1.085	-0.041	1.079
13	-0.035	1.076	-0.072	1.075	-0.013	1.059	-0.035	1.074
14	-0.033	1.070	-0.028	1.068	-0.040	1.075	-0.055	1.085
15	-0.020	1.060	-0.030	1.065	-0.036	1.065	-0.025	1.064
16	-0.023	1.061	-0.023	1.061	-0.039	1.061	-0.015	1.056
17	-0.017	1.057	-0.016	1.056	-0.015	1.055	-0.012	1.053
18	-0.029	1.063	-0.018	1.054	-0.019	1.056	-0.014	1.053
19	-0.046	1.048	-0.030	1.062	0.001	1.044	-0.013	1.052
20	-0.002	1.044	-0.029	1.063	-0.050	1.049	0.029	1.025
21	-0.007	1.047	-0.012	1.048	-0.012	1.050	0.027	1.027
22	-0.008	1.046	-0.004	1.044	-0.009	1.046	-0.004	1.044
23	-0.004	1.044	-0.004	1.044	-0.006	1.045	-0.002	1.042
24	-0.005	1.043	-0.007	1.043	-0.017	1.042	0.000	1.040
25	-0.007	1.044	-0.001	1.040	-0.009	1.041	0.000	1.039
26	-0.003	1.041	-0.022	1.039	-0.003	1.040	-0.000	1.039
27	0.001	1.039	-0.003	1.039	-0.007	1.043	0.005	1.035
28	0.000	1.038	0.004	1.036	-0.004	1.038	-0.000	1.038
29	0.002	1.037	0.000	1.038	0.002	1.036	0.003	1.036
30	0.002	1.036	-0.000	1.036	-0.011	1.035	0.004	1.035
31	0.004	1.035	0.005	1.034	0.007	1.033	0.004	1.034
32	0.003	1.036	0.005	1.034	-0.007	1.035	0.006	1.033
33	0.007	1.033	0.002	1.034	0.005	1.033	0.004	1.034
34	0.005	1.034	0.007	1.032	0.005	1.033	0.007	1.032
35	0.011	1.031	0.008	1.032	0.001	1.032	0.008	1.031
36	0.018	1.026	0.006	1.032	0.003	1.032	0.008	1.031
37	-0.030	1.056	0.020	1.025	0.010	1.030	0.008	1.031
38	-0.043	1.032	-0.008	1.043	0.003	1.035	0.009	1.030
39			-0.040	1.030	0.066	0.996	-0.003	1.039
40					-0.038	1.028	0.054	1.003
41							-0.035	1.027

	42		43		44		45	
5	-0.219	1.231	-0.227	1.239	-0.214	1.232	-0.293	1.272
6	-0.192	1.192	-0.157	1.178	-0.145	1.173	-0.191	1.201
7	-0.162	1.161	-0.112	1.141	-0.111	1.143	-0.122	1.151
8	-0.080	1.117	-0.102	1.131	-0.115	1.135	-0.095	1.127
9	-0.086	1.114	-0.084	1.115	-0.100	1.127	-0.111	1.116
10	-0.058	1.095	-0.057	1.095	-0.064	1.098	-0.079	1.102
11	-0.051	1.087	-0.048	1.088	-0.079	1.087	-0.056	1.089
12	-0.056	1.080	-0.039	1.077	-0.053	1.082	-0.050	1.082
13	-0.031	1.071	-0.032	1.072	-0.032	1.071	-0.038	1.075
14	-0.061	1.066	-0.029	1.066	-0.032	1.069	-0.029	1.067
15	-0.031	1.065	-0.032	1.068	-0.017	1.059	-0.052	1.060
16	-0.024	1.061	-0.018	1.057	-0.029	1.061	-0.018	1.057
17	-0.023	1.060	-0.014	1.053	-0.015	1.054	-0.017	1.056
18	-0.025	1.053	-0.015	1.054	-0.015	1.053	-0.032	1.053
19	-0.009	1.049	-0.009	1.048	-0.013	1.052	-0.011	1.050
20	-0.022	1.056	-0.008	1.047	-0.015	1.048	-0.017	1.048
21	-0.042	1.043	-0.028	1.058	0.002	1.041	-0.010	1.046
22	0.000	1.041	-0.031	1.061	-0.045	1.045	0.002	1.038
23	-0.004	1.043	0.000	1.040	-0.004	1.043	0.022	1.027
24	-0.012	1.042	-0.001	1.040	-0.006	1.040	-0.007	1.042
25	0.004	1.037	-0.001	1.039	0.001	1.038	-0.007	1.039
26	-0.003	1.040	0.002	1.037	-0.002	1.039	0.001	1.037
27	-0.002	1.038	0.001	1.037	0.002	1.036	-0.012	1.037
28	-0.018	1.035	0.004	1.035	-0.003	1.036	0.003	1.035
29	0.004	1.035	0.008	1.032	0.005	1.033	0.004	1.034
30	-0.004	1.035	0.005	1.034	0.003	1.034	-0.017	1.034
31	0.004	1.034	0.006	1.033	0.005	1.033	0.005	1.033
32	0.004	1.033	0.007	1.031	0.001	1.033	0.006	1.032
33	0.003	1.033	0.006	1.032	-0.009	1.032	0.003	1.032
34	0.005	1.032	0.006	1.032	0.005	1.032	0.004	1.033
35	-0.001	1.031	0.008	1.031	0.011	1.028	0.001	1.031
36	0.001	1.031	0.007	1.031	0.004	1.030	-0.003	1.030
37	0.010	1.029	0.008	1.030	0.009	1.029	0.009	1.029
38	0.008	1.030	0.010	1.029	0.007	1.030	0.009	1.028
39	0.006	1.030	0.010	1.028	0.010	1.028	0.007	1.029
40	0.006	1.032	0.009	1.029	0.006	1.029	0.005	1.028
41	-0.006	1.040	0.017	1.025	0.012	1.027	0.010	1.028
42	-0.033	1.025	-0.029	1.053	0.020	1.022	0.011	1.026
43			-0.035	1.026	-0.014	1.043	0.013	1.026
44					-0.036	1.027	0.028	1.016
45							-0.034	1.026

	46		47		48		49	
5	-0.241	1.244	-0.223	1.235	-0.233	1.242	-0.237	1.247
6	-0.180	1.196	-0.180	1.198	-0.211	1.205	-0.163	1.183
7	-0.125	1.151	-0.126	1.153	-0.121	1.152	-0.150	1.156
8	-0.109	1.136	-0.100	1.132	-0.118	1.127	-0.094	1.124
9	-0.075	1.107	-0.067	1.104	-0.090	1.116	-0.081	1.114
10	-0.062	1.097	-0.055	1.092	-0.069	1.102	-0.073	1.106
11	-0.046	1.084	-0.057	1.091	-0.048	1.085	-0.060	1.093
12	-0.049	1.083	-0.021	1.068	-0.068	1.078	-0.047	1.080
13	-0.032	1.070	-0.034	1.072	-0.033	1.071	-0.032	1.070
14	-0.033	1.069	-0.028	1.066	-0.034	1.069	-0.042	1.067
15	-0.034	1.067	-0.017	1.058	-0.030	1.064	-0.024	1.061
16	-0.022	1.060	-0.008	1.051	-0.051	1.057	-0.010	1.051
17	-0.015	1.054	-0.015	1.054	-0.016	1.054	-0.013	1.052
18	-0.019	1.055	-0.013	1.051	-0.023	1.052	-0.012	1.051
19	-0.010	1.049	-0.010	1.049	-0.013	1.050	-0.011	1.049
20	-0.011	1.048	-0.005	1.045	-0.015	1.048	-0.010	1.047
21	-0.008	1.046	-0.007	1.045	-0.011	1.046	-0.017	1.044
22	-0.013	1.048	-0.001	1.041	-0.007	1.044	-0.003	1.042
23	-0.039	1.040	0.008	1.035	-0.000	1.040	-0.004	1.042
24	0.000	1.039	-0.008	1.045	-0.037	1.039	-0.003	1.040
25	0.001	1.038	0.001	1.038	-0.001	1.039	0.008	1.033
26	-0.001	1.038	0.000	1.037	-0.000	1.038	0.004	1.035
27	0.001	1.037	-0.001	1.038	-0.002	1.037	0.002	1.036
28	-0.001	1.037	0.002	1.035	-0.002	1.035	-0.006	1.035
29	0.004	1.034	0.004	1.034	0.004	1.034	0.004	1.034
30	0.001	1.035	0.005	1.033	-0.004	1.034	0.005	1.032
31	-0.003	1.038	0.001	1.035	0.004	1.033	0.006	1.032
32	0.003	1.033	0.006	1.032	-0.014	1.031	0.004	1.032
33	0.007	1.031	0.007	1.031	0.004	1.031	0.001	1.034
34	0.005	1.031	0.007	1.030	0.005	1.031	0.006	1.030
35	0.007	1.031	0.005	1.031	0.008	1.029	-0.000	1.030
36	0.006	1.030	0.009	1.029	-0.006	1.029	0.007	1.029
37	0.008	1.029	0.009	1.029	0.009	1.028	0.009	1.028
38	0.007	1.029	0.010	1.028	0.007	1.029	0.010	1.027
39	0.010	1.028	0.009	1.028	0.007	1.028	0.011	1.026
40	0.009	1.028	0.010	1.027	0.001	1.027	0.011	1.026
41	0.010	1.027	0.010	1.027	0.011	1.026	0.010	1.027
42	0.010	1.027	0.011	1.026	0.005	1.027	0.004	1.026
43	0.015	1.025	0.013	1.025	0.012	1.026	0.011	1.026
44	0.002	1.032	0.013	1.025	0.009	1.026	0.012	1.025
45	0.040	1.009	0.009	1.028	0.008	1.027	0.012	1.025
46	-0.032	1.024	0.021	1.022	0.016	1.023	0.011	1.026
47			-0.030	1.023	0.016	1.025	0.018	1.022
48					-0.032	1.024	-0.002	1.034
49							-0.030	1.023

50

5	-0.287	1.269
6	-0.154	1.179
7	-0.120	1.146
8	-0.088	1.122
9	-0.077	1.110
10	-0.094	1.102
11	-0.046	1.083
12	-0.042	1.079
13	-0.041	1.076
14	-0.032	1.068
15	-0.034	1.063
16	-0.022	1.058
17	-0.034	1.066
18	-0.016	1.053
19	-0.008	1.047
20	-0.025	1.047
21	-0.007	1.045
22	-0.006	1.043
23	-0.001	1.040
24	-0.006	1.043
25	-0.036	1.038
26	0.000	1.037
27	0.001	1.036
28	0.001	1.035
29	0.003	1.034
30	-0.009	1.033
31	0.005	1.032
32	0.004	1.032
33	0.006	1.031
34	0.003	1.032
35	0.003	1.029
36	0.006	1.029
37	0.009	1.028
38	0.008	1.028
39	0.009	1.027
40	-0.001	1.027
41	0.011	1.026
42	0.010	1.026
43	0.012	1.025
44	0.011	1.025
45	0.008	1.025
46	0.011	1.025
47	0.014	1.024
48	0.012	1.025
49	0.019	1.020
50	-0.029	1.022

CHAPTER SIX

UNDERSTANDING WHY THE CURVE METHOD WORKS

The k-sample p-value curves presented in Chapter Two imply that the k-sample p-value can be approximated by a simple function of the two-sample p-value. Indeed, for sample sizes under 100, we can directly observe that at conventional significance levels, the three-sample p-value is almost exactly a function of the two-sample p-value. We would like to investigate why this, in fact, occurs.

Let U(2) signify the properly normalized two sample statistic and U(k) represent the k sample statistic. U(2) has asymptotic distribution $F_2(t)$ which is well known, continuous. We assume that the unknown asymptotic distibution of U(k), $F_k(t)$, is also continuous. Let $G_2(t) = 1 - F_2(t)$ and $G_k(t) = 1 - F_k(t)$. Thus, for large samples we know:

$$\gamma \approx G_2(t) \text{ and } \alpha_k \approx G_k(t).$$

Therefore, in large samples, $\alpha_k \approx G_k(G_2^{-1}(\gamma))$. In other words, α_k is some smooth function of γ.

However, this does not explain why the function $G_k o G_2^{-1}$ appears to work so well for small samples. This is especially surprising since $G_2(t)$ itself is a poor formula for p-values in small samples. To understand this phenomenon, let us focus attention on the three equal sample sizes case.

We shall start with the premise that for every n there exist two functions such that $\gamma = G_{n2}(t)$ and $\alpha = G_{n3}(t)$. We know that $G_2(t)$ is not an especially good approximation for $G_{n2}(t)$ for small n. However, it might be reasonable to assume that the two can be brought into better agreement with some simple adjustment; after all, one commonly uses a continuity correction when approximating the binomial distribution with the (asymptotically correct) normal distribution. More generally one could consider a translation and a scale change, i.e., a simple linear transformation (as discussed in Chapter Five). Returning to our premise, we know that $\alpha = G_{n3}(G_{n2}^{-1}(\gamma))$. We would like to know how this is related to $G_3(G_2^{-1}(\gamma))$, under the assumption that a linear transformation makes the asymptotic formula accurate. We assume that:

$$\gamma = G_{n2}(t) \approx G_2(a_{n2} + b_{n2}t)$$

$$\alpha = G_{n3}(t) \approx G_3(a_{n3} + b_{n3}t)$$

Since $(a_{n2}+b_{n2}t) \approx G_2^{-1}(\gamma)$, we see that $t \approx [G_2^{-1}(\gamma)-a_{n2}]/b_{n2}$ and we also know that $t=G_{n2}^{-1}(\gamma)$. Therefore,

$$\alpha = G_{n3}(G_{n2}^{-1}(\gamma))$$
$$\approx G_3(a_{n3} + b_{n3}[(G_2^{-1}(\gamma)-a_{n2})/b_{n2}])$$
$$\approx G_3[a_{n3} - (b_{n3}/b_{n2})a_{n2} + (b_{n3}/b_{n2})G_2^{-1}(\gamma)].$$

Further, if $a_{n2} = a_{n3}$ and $b_{n2} = b_{n3}$, then simple algebra dictates that $\alpha \approx G_3(G_2^{-1}(\gamma))$. Thus, if the location and scale shifts are approximately the same for both two and three samples, for small n, then $G_3(G_2^{-1}(\gamma))$ is a proper function for even small n.

Now let us determine the validity of this linear transformation hypothesis. When considering $\gamma < .05$, $G_2(t) = 2\exp(-2t^2)$ almost exactly. This is because the remaining terms in the series of the asymptotic formula are negligible. Using the curve for k=3 in Chapter Two and inserting the asymptotic approximation for γ, we can determine a reasonable approximate asymptotic formula for three samples. Incidentally, although we know that this formula provides good asymptotic three-sample p-values for n between 50 and 100, we can only assume that this relationship still holds as n goes to infinity. Thus we can estimate the large sample function of t as follows:

$$G_3(t) = 3[2\exp(-2t^2)] - 1.5735[2\exp(-2t^2)]1.3916$$

For small n, it is possible to determine an adjusted value t* such that $G_2(t^*) = \gamma$ exactly. For t > 1.40, we regressed this adjusted value of t* on the critical value t for the two sample case obtaining a_{n2} and b_{n2}, for each n from 5 to 50. We did likewise to obtain a_{n3} and b_{n3} for the three sample case. All regressions yielded $R^2 > .999$. Consequently, we could conclude that $G_2(a_{n2}+b_{n2}t)$ and $G_3(a_{n3}+b_{n3}t)$ were almost exact estimates of γ and α. In other words, the true function of t, needs both a location and scale shift to lie on the asymptotic curve, as we hypothesized. In addition, for all n examined, $a_{n2} \approx a_{n3}$ and $b_{n2} \approx b_{n3}$.

For example, for n=10, we observed for both two and three samples, that $a \approx -.2228$ and $b \approx 1.1613$. For t=1.5652, we know $\gamma=.0123$ and $\alpha=.0336$. The asymptotic p-values are as follows:

$$G_2(t) = .0149$$

$$G_2(a+bt) = .0123$$

$$G_3(t) = .0402$$

$$G_3(a+bt) = .0336$$

Now, for t=1.7889 ($\gamma=.00206$ and $\alpha=.0059$), the asymptotic pvalues are as follows:

$$G_2(t) = .0033$$

$$G_2(a+bt) = .00206$$

$$G_3(t) = .0093$$

$$G_3(a+bt) = .0059.$$

The following presents the estimates of a and b found for various values of n. For a given n, the estimates vary only slightly between the two and three sample cases.

n	a_{n2}	b_{n2}	a_{n3}	b_{n3}
6	-.3966	1.2958	-.3934	1.2942
10	-.2229	1.1614	-.2227	1.1613
20	-.0747	1.0569	-.0728	1.0557
30	-.0509	1.0384	-.0494	1.0376
40	-.0378	1.0285	-.0367	1.0278
50	-.0286	1.0218	-.0269	1.0207

In conclusion, for a given n, the linear transformation of t required to make the two-sample asymptotic formula accurate, is almost exactly the same as that required to transform t to make the three-sample approximate asymptotic formula work. We demonstrated earlier, that under this circumstance, the composition of asymptotic functions would equal the composition of the exact formulas.

In addition, if we are uncomfortable ignoring the slight deviations from equality, when we consider the actual values obtained for a_{n2}, a_{n3}, b_{n2}, and b_{n3}, the discrepancies have essentially no impact. For example, in the case of n=40,

$$\alpha = G_3(-.0367 + (.0378) + (.9993)G_2^{-1}(\gamma)$$

$$= G_3(.0011 + (.9993)G_2^{-1}(\gamma))$$

$$\approx G_3(G_2^{-1}(\gamma)), \text{ when } \gamma \text{ is small.}$$

CONCLUSION

It is quite clear that in large samples, α_k is a function of γ. We now see that the location and scale deviation from the asymptotic function is almost equal for the two and three sample case for a given n. This phenomenon is sufficient to explain why $G_3 o G_2^{-1}$ is an accurate function for sample sizes as small as five. We presume that the same relationship exists for k in general.

In other words, for each value of n, there is an adjustment of t necessary before the asymptotic formula can yield an accurate p-value; this adjustment is approximately the same for any number of samples.

What still remains to be explained is precisely why the linear transformation for the two-sample case is the same as that required for the three-sample case.

C H A P T E R S E V E N

AN APPLICATION

In this chapter, the k-sample Kolmogorov-Smirnov procedure is applied to a set of data from the Cincinnati Clinic of the Lipid Research Clinics Program. The data are triglyceride and cholesterol measures from children aged 10 and 11. The procedure is used to compare the empirical distribution of triglyceride and then of cholesterol in the four race-sex groups: White Females, Black Females, White Males, and Black Males. The respective sample sizes are: 71, 22, 61, and 20. A computer program in PASCAL which computes U for each pair of samples, and then the procedure-wise estimated p-value using the curves, appears in Appendix 7.1.

TRIGLYCERIDES

When triglycerides is the variable of interest, we observe the following two-sample statistics and p-values:

	U(2)	p-value
White Females vs. Black Females	1.062	.1716
White Females vs. White Males	1.614	.0083
White Females vs. Black Males	1.221	.0773
Black Females vs. White Males	.569	.8476
Black Females vs. Black Males	.588	.7928
White Males vs. Black Males	.728	.5928

Thus, the k-sample statistic is 1.614. The two-sample p-values presented in the above table are not needed in the calculation of the four-sample p-value. They are presented simply to convey some information about each paired comparison. These p-values should not be taken seriously due to the number of comparisons tested. The two-sample p-values associated with U(k) and each pair's sample sizes are as follows:

	U(k)	p-value
White Females vs. Black Females	1.614	.0071
White Females vs. White Males	1.614	.0083
White Females vs. Black Males	1.614	.0071
Black Females vs. White Males	1.614	.0071
Black Females vs. Black Males	1.614	.0059
White Males vs. Black Males	1.614	.0074

The average of these p-values equals .0072. When this is plugged into the four sample curve function, we obtain the procedure-wise p-value: .0369. Therefore we can infer that the four race-sex groups are not homogeneous with respect to the distribution of triglycerides in children aged 10-11.

CHOLESTEROL

Now, let us test the null hypothesis that the four race-sex groups have the same distribution of cholesterol among 10-11 year olds. (Note: due to missing values, the white males group is now 58 in number, and black females are 20 in number.) When cholesterol is the variable of interest, we observe the following two-sample statistics and p-values:

	U(2)	p-value
White Females vs. Black Females	.439	.9747
White Females vs. White Males	.599	.8127
White Females vs. Black Males	.859	.3847
Black Females vs. White Males	.625	.7608
Black Females vs. Black Males	.790	.5713
White Males vs. Black Males	.890	.3408

Thus, the k-sample statistic is .890. The two-sample p-values associated with U(k) and each pair's sample sizes are as follows:

	U(k)	p-value
White Females vs. Black Females	.890	.3407
White Females vs. White Males	.890	.3539
White Females vs. Black Males	.890	.3407
Black Females vs. White Males	.890	.3310
Black Females vs. Black Males	.890	.3356
White Males vs. Black Males	.890	.3408

The average of these p-values equals .3404. When this is plugged into the four sample curve function, we obtain the procedure-wise p-value: .8214. Therefore we cannot reject the null hypothesis that the four race-sex groups are homogeneous with respect to the distribution of cholesterol in children aged 10-11.

Appendix 7.1

This program uses the curve method to determine p-values for the k-sample Kolmogorov-Smirnov procedure. At the beginning of the program the user must supply information, as described in the comments. It is currently set up to do the triglyceride analysis. At the end of the program, each function or procedure is described. It is designed to be used in parts, if desired. A computational technique suggested by Kim and Jennrich (1970) to keep the numbers small in the two-sample p-value evaluation is used.

The program is written as a series of subroutines, so that the user who only needs a program which combines p-values associated with the maximum two-sample U statistic for example, can just use function "pvaluek". Or if a user already knows the maximum two-sample value of the statistic U, he or she only need use: function "combin", function "pathpvalue", function "pbar" and function "pvaluek".

```
program ksample;

(* user must set up these constants:
      ntotal = number of observations in combined samples
      inputdata = data file
      maxk = number of samples
      n1 = a number greater than or equal to the number in
Sample 1
      n2 = a number greater than or equal to the number in
Sample 2, etc.,
              one needs to specify as many n's as there are
samples
```

```
      nmax = a number greater than or equal to the
             number of observations in largest sample *)

const ntotal=174; inputdata = 'KSDATA.'; maxk = 4;
      n1 = 200; n2 =100; n3 =200; n4=100;  nmax=200;

type bigreal = array[1..ntotal] of real;
     bigint = array[1..ntotal] of integer;
     doub = array[0..ntotal] of array[1..10] of real;
     smallint = array[1..10] of integer;
     smallrel = array[1..10] of array[1,.10] of real;

function combin(b,nna:integer):real;
var z:real;
var j:integer;
begin
if b=0 then combin:=1.0;
if b>0 then begin
z:=1.0;
for j:=nna+1 to nna+b do begin;
    z:=((j-nna)/j)*z;
combin:=z; end;
end;
end;

function pathpvalue(u:real;nna,nnb:integer):real;
var pvalue,a1,x,w:real;
var j,a,b,thisa,lasta,holda:integer;
var h:array [0..1] of array [-1..nmax] of real;

begin
a1:=1.0;
w:=u/sqrt((1.0*nna*nnb)/(nna+nnb));

for a:=0 to 1 do begin for b:=-1 to nnb do begin
h[a,b]:=0.0;
end;end;

thisa:=0; lasta:=1;

for a := 0 to nna do begin
if a = 0 then h[0,0]:=a1;
holda:=thisa; thisa:=lasta; lasta:=holda;

for b:=0 to nnb do begin

if a*b=0 then begin
   if abs(a*nnb-b*nna)>=w*nna*nnb then h[thisa,b]:=0.0
   else h[thisa,b]:=combin(b,nna);
end;

if a*b>0 then begin
if abs(a*nnb-b*nna)>=w*nna*nnb then h[thisa,b]:=0.0
   else h[thisa,b]:=(h[lasta,b]+(h[thisa,b-1]*b/(b+nna)));
```

```
end;end;end;
pathpvalue := 1 - h[thisa,nnb];
writeln(lst,u:10:6, 1-h[thisa,nnb]:10:6, nna:6, nnb:6);
end;

procedure sort(var x:bigreal;ntot:integer;
               var y:bigint);
var g,i,j,j1,tempi : integer;
var temp : real;
begin;
g:=ntot div 2;
while g>0 do begin;
for i:=1 to ntot-g do begin; j:=i;
while j>0 do begin;
j1:=j+g;
if x[j]>x[j1] then begin;
temp:=x[j]; x[j]:=x[j1]; x[j1]:=temp;
tempi:=y[j]; y[j]:=y[j1]; y[j1]:=tempi;
j:=j-g;
end (if x[j])
else j:=0
end (while j);
end (for i);
g:= g div 2;
end (while g);
end;

procedure collapse(var k,ntot:integer;var x:bigreal;
                  var y:bigint; var samp:doub);
var i,j,l:integer;
begin;
  for i:=1 to ntot do begin
    for j:=1 to k do begin
      if y[i]=j then samp[i,j]:=1.0  else samp[i,j] := 0.0;
    end;
  end;
  writeln(lst);

  i := 1;
  while i < ntot do begin;
    if x[i] < x[i+1] then i := i + 1
    else begin;
      for j := 1 to k do samp[i,j] := samp[i,j] +
samp[i+1,j];
      if i < ntot-1 then  for l := i+1 to ntot-1 do begin;
        x[l] := x[l+1];
        for j := 1 to k do samp[l,j] := samp[l+1,j];
      end;
      ntot := ntot-1;
    end;
  end;
end;

function statistic(k,ntot:integer;var samp:doub;var
```

```
n:smallint;
                   var x:bigreal):real;
var prop,sum:doub;
var u: array[1..10] of array [1..10] of real;
var pvalue,umax:real;
var i,j,l:integer;
begin;
   for i := 0 to ntot do for j := 1 to k do sum[i,j] := 0.0;
   writeln(lst,'CDF for each sample at each unique
observation');
   writeln(lst);
   for i:=1 to ntot do begin
      write(lst,x[i]:8:2);
   for j:=1 to k do begin
      sum[i,j]:=sum[i-1,j]+samp[i,j];
      prop[i,j]:=sum[i,j]/n[j];
      write(lst,prop[i,j]:8:4);
   end;
   writeln(lst);
   end;
  writeln(lst);

   for i := 1 to k do for j := 1 to k do u[i,j] := 0.0;
   for l:=1 to ntot do begin
   for i:=1 to k-1 do begin
   for j:=i+1 to k do begin
      if abs(prop[l,i]-prop[l,j])>u[i,j]
         then u[i,j]:=abs(prop[l,i]-prop[l,j]);
   end;
   end;
   end;
   umax:=0;
   writeln(lst,'samp i samp j      u       pvalue    n[i]
n[j]');
   writeln(lst);
   for i:=1 to k-1 do begin;
   for j:=i+1 to k do begin;
   u[i,j]:=u[i,j]*sqrt((n[i]*n[j])/(n[i]+n[j]));
   write(lst,i:6,j:6);
   pvalue:=pathpvalue(u[i,j],n[i],n[j]);
   if u[i,j]>umax then umax:=u[i,j];
   end;
   end;
writeln(lst);
statistic:=umax;
end;

function pbar(k:integer;umax:real;var n:smallint):real;
  var sum:real;
  var i,j,kchoose2:integer;
  var pvalue:smallrel;
begin;
sum:=0.0; kchoose2:=0;
for i:=1 to k-1 do begin for j:=i+1 to k do begin
```

```
write(lst,i:6,j:6);
pvalue[i,j]:=pathpvalue(umax,n[i],n[j]);
sum:=pvalue[i,j]+sum;
kchoose2:=kchoose2+1;
end;end; writeln(lst); writeln(lst,'pbar =
',sum/kchoose2:10:6);
pbar:=sum/kchoose2;
end;

function pvaluek(k:integer;pbar:real):real; label return;
begin;
if k=2 then begin;
  pvaluek:=pbar; goto return; end;
if k=3 then begin;
  pvaluek:=3*pbar - 1.5735*(exp(ln(pbar)*1.3916)); goto
return; end;
if k=4 then begin;
  pvaluek:=6*pbar - 5.3761*(exp(ln(pbar)*1.3755)); goto
return; end;
if k=5 then begin;
  pvaluek:=10*pbar - 11.4256*(exp(ln(pbar)*1.3594)); goto
return; end;
if k=6 then begin;
  pvaluek:=15*pbar - 19.3440*(exp(ln(pbar)*1.3431)); goto
return; end;
if k=7 then begin;
  pvaluek:=21*pbar - 28.4718*(exp(ln(pbar)*1.3263)); goto
return; end;
if k=8 then begin;
  pvaluek:=28*pbar - 37.5653*(exp(ln(pbar)*1.3073)); goto
return; end;
if k=9 then begin;
  pvaluek:=36*pbar - 47.4433*(exp(ln(pbar)*1.2913)); goto
return; end;
if k=10 then begin;
  pvaluek:=45*pbar - 54.3065*(exp(ln(pbar)*1.2693)); goto
return; end;
pvaluek:=1.0;
return:
end;

var x:bigreal;
var y:bigint;
var i,j,k,ntot:integer;
var pvalue,umax,junk:real;
var samp:doub;
var n:smallint;
var file1:text;

begin;
assign(file1,inputdata); reset(file1);
ntot:=ntotal;
for i:=1 to ntot do begin;
readln(file1,y[i],x[i]);
```

```
end;

k:=maxk;

for j:=1 to k  do n[j]:=0;

for i:=1 to ntot do begin
for j:=1 to k do begin
   if y[i]=j then n[j]:=n[j]+1;
end;
end;

(* Procedure sort orders all of the observations from small
to big *)

sort(x,ntot,y);

(* Procedure collapse finds the number of observations from
each sample for
   each unique observation *)

collapse(k,ntot,x,y,samp);

(* Function statistic yields the two-sample statistics and
corresponding
   p-values.  It also produces the k-sample statistic.  It
call function
   pathpvalue to get the two-sample p-values, which in turn,
calls
   function combin to compute needed combinatoric
information *)

umax:=statistic(k,ntot,samp,n,x);

(* Function pvaluek calls pbar which computes the average
two-sample
   p-value corresponding to the k-sample statistic.  Pbar
utilizes pathpvalue
   to get the p-values.  Function pvaluek plugs the average
p-value
   into the relevant curve equation to yield the k-sample
p-value. *)

pvalue:=pvaluek(k,pbar(k,umax,n)); writeln(lst,'pvaluek =
',pvalue:10:6);

 if pvalue>0.1 then
     writeln(lst,'WARNING: P-value>.10: method may be
inaccurate');
 if (pvalue>0.05) and (k>7) then
     writeln(lst,'WARNING: P-value>.05 and k>7: Result is
quite conservative');

end.
```

CHAPTER EIGHT

FUTURE WORK

This research can be extended in several directions. Of particular importance are the following: 1) extension to censored data, 2) comparison of this procedure with others previously proposed with respect to size and power, and 3) investigation of the generalizability of the approach to other k-sample settings.

Perhaps the most useful extension is to censored data. That is, we could develop the ability to use a k-sample Kolmogorov-Smirnov approach to compare survival in situations where some subjects do not fail. As a first attempt, one could translate two-sample p-values from the Fleming et al (1980) modified Kolmogorov-Smirnov approach to k-sample p-values using the already established curves. The validity of the method could be tested with simulation techniques under a variety of conditions.

Another matter of interest is comparison of the procedure with those developed by Kiefer and by Conover. Again, a simulation study could be conducted which compares the size under the null hypothesis and the relative power of the three procedures under a variety of circumstances.

The third area of interest centers around the issue of generalizability to other k-sample settings or, more generally, other multiple comparisons settings. One could address the question of whether any other distributions for k-sample statistics are functions of the distributions of two-sample statistics. In addition, one could explore the adequacy of the $S_1 - S_2 + S_3$ approximation for a variety of situations. If this proves to be a good estimate in general, possibly methods could be developed to estimate S_2 and S_3 in general.

B I B L I O G R A P H Y

Birnbaum ZW and Hall RA. Small sample distributions for multi-sample statistics of the Smirnov Type. Annals of Mathematical Statistics 31: 710-720. 1960.

Birnbaum ZW and Tingey FH. One sided confidence contours for probability distribution functions. Annals of Mathematical Statistics 22: 592-596. 1951.

Conover WJ. Several k-sample Kolmogorov-Smirnov tests. Annals of Mathematical Statistics 36: 1019-1026. 1965.

Conover WJ. A k-sample extension of the one-sided two-sample Smirnov test statistic. Annals of Mathematical Statistics 38: 1726-1730. 1967.

Darling DA. The Kolmogorov-Smirnov, Cramer-Von Mises tests. Annals of Mathematical Statistics 28: 823-838. 1957.

David HT. A three-sample Kolmogorov-Smirnov test. Annals of Mathematical Statistics 29: 842-851. 1958.

Feller W. An Introduction to Probability Theory and its Applications. Volume 1, third edition, John Wiley, New York, 1968.

Fleming TR, O'Fallon JR, O'Brien PC, Harrington DP. Modified Kolmogorov-Smirnov test procedures with application to arbitrarily right censored data. Biometrica 36: 607-625, 1980.

Gardner RH, Pinder JE, Wood RS. Monte Carlo estimation of percentiles for the multi-Smirnov test. Journal of Statistical Computation and Simulation 10: 243-249. 1980.

Gnedenko BV and Korolyuk VC. On the maximal deviation between two empirical distributions. Doklady Akademii Nauk SSSR 80: 525-528. 1951.

Harter HL. Modified asymptotic formulas for critical values of the Kolmogorov test statistic. The American Statistician 34: 110-111. 1980.

Hodges JL. The significance probability of the Smirnov two sample test. Arkiv for Matematik 3: 469-486, 1957.

Jonckheere AR. A distribution free k-sample test against ordered alternatives. Biometrika 41: 133-145. 1954.

Kiefer J. K-sample analogues of the Kolmogorov-Smirnov and Cramer-Von Mises tests. Annals of Mathematical Statistics 30: 420-447. 1959.

Kim PJ. On the exact and approximate sampling distribution of the two sample Kolmogorov-Smirnov criterion Dmn, m≤n. Journal of the American Statistical Association 64: 1625-1637. 1969.

Kim PJ and Jennrich RI. Selected Tables in Mathematical Statistics I (Table 2.1) Editors: HL Harter and DB Owen. Markham, Chicago, 1970.

Kolmogorov AN. Sulla determinazione empirica di une legge di distribuzione. Giornale dell'Istituto Italiano degli Attuari 4: 83-91. 1933.

Korolyuk VC. On the deviation of empirical distributions for the case of two independent samples. Izvestii Akademii Nauk SSSR, Seriya Matematiki 19: 81-96. 1955.

Massey FJ. The distribution of the maximum deviation between two sample cumulative step functions Annals of Mathematical Statistics 22: 125-128. 1951.

Massey FJ. Distribution table for the deviation between two sample cumulatives. Annals of Mathematical Statistics 23: 4435-441. 1952.

Smirnov NV. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bulletin Mathematique de l'Universite de Moscou, Serie Internationale 2: 1-6. 1939.

Smirnov NV. Table for estimating goodness of fit of empirical distributions. Annals of Mathematical Statistics 19: 279-281. 1948.

Steck GP. The Smirnov two sample tests as rank tests. Annals of Mathematical Statistics 40: 1449-1466. 1969.

Taylor MS and Becker MP. Quantiles of multi-sample Smirnov type statistics. Journal of Statistical Computation and Simulation 16: 25-34. 1982.

Wallenstein S. Distributions of some one-sided k-sample Smirnov-type statistics. Journal American Statistical Association 75: 441-446. 1980.

Wolf EH and Naus JI. Tables of critical values for a k-sample Kolmogorov-Smirnov test statistic. Journal of the

American Statistical Association 68: 994-997. 1973.