AD-P003 844

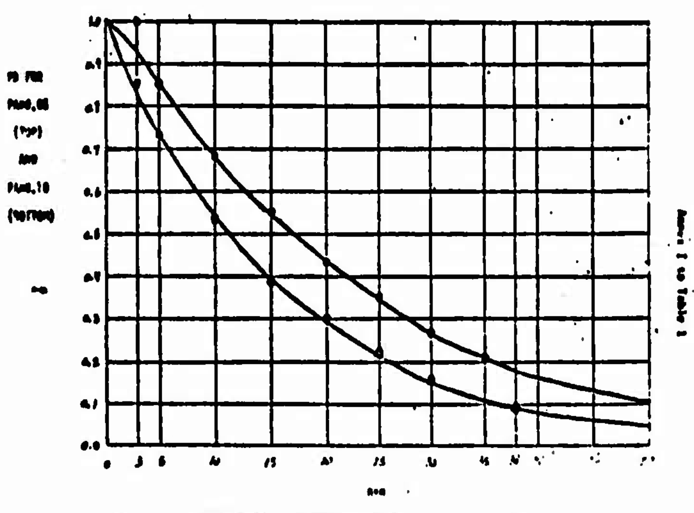
ON THE LEHMANN POWER ANALYSIS FOR THE WILCOXON RANK SUM TEST

James R. Knaub, Jr.

US Army Logistics Center

ABSTRACT

The Wilcoxon Rank Sum (or Mann-Whitney) Test is among the most useful and powerful of the non-parametric hypothesis tests. However, as with many hypothesis tests, when a clear alternative hypothesis and corresponding power analysis is not present, the practical interpretation of results using this test suffers greatly. This paper presents and clarifies an alternative suggestd by E. L. Lehmann in 1953 and provides tables of practical use which have not prviously been calculated due to computational difficulties.



POINTS GRAPHED ARE FROM SIMULATION
TOP CURVE IS $f_{0.05}(x) =$ [illegible]
BOTTOM CURVE IS $f_{0.10}(x) = EXP(-X/16)$

On the Lehmann Power Analysis for the Wilcoxon Rank Sum Test

The Wilcoxon Rank Sum (or Mann-Whitney) Test is among the most useful and powerful of the non-parametric hypothesis tests. However, as with many hypothesis tests, when a clear alternative hypothesis and corresponding power analysis is not present, the practical interpretation of results using this test suffers greatly. This paper presents and clarifies an alternative suggested by E. L. Lehmann in 1953 (Annals of Mathematical Statistics [7]) and provides tables of practical use which have not previously been calculated due to computational difficulties. This work has recently been applied to survey data gathered for the US Army Logistics Center. (See reference [5].)

When sample sizes are small, and a power analysis is not available, one may fail to reject the null hypothesis when the true state of nature is very different from what is stated in the null hypothesis. With a small sample size and small α, it may be impossible to reject H_0. Further, when sample sizes are very large, the null hypothesis may be rejected at a very small significance level when actually the null hypothesis is so nearly true, that it is close enough for all practical purposes. Taken to the extreme, with infinite sample sizes, the attained significance level will be zero, even when there is only a very small, but finite difference between H_0 and the true state of nature. Thus significance level can be very misleading if used alone.

When a null and a definitive alternative hypothesis can both be stated, and probability distributions found under each, the results of an hypothesis test can be stated similarly to a confidence interval if the "point estimate" from the observed values falls between the two hypotheses. In the case of the Wilcoxon Rank Sum Test, only one alternative hypothesis has been well developed and will be presented here. Due to the nature of this test, however, even if the evidence may strongly indicate that the true state of nature is not bounded between this alternative and the null hypothesis, this power analysis can still be used to obtain a reasonable estimate of what the actual state of nature happens to be. (In the case of the Multiple-sample Westenberg-type tests of reference [4], an alternative must be picked such that the true state of nature is indicated to be bounded by the null and alternative hypotheses. Fortunately, that is not the case here, nor was it the case in reference [6], which is a multi-sample test.)

Consider that the null hypothesis, H_0, of the Wilcoxon Rank Sum Test indicates that $P(X<Y) = 1/2$. That is, under H_0, any value picked at random from the Y population, is larger than any value picked at random from the X population, with probability of 1/2. Here an alternative hypothesis, H_1, is used such that $P(X<Y) = 2/3$. (The exact form of H_1 is discussed in [7].) Graph 1 illustrates a possible configuration for this alternative hypothesis. For this example, consider that under H_0, all observations are taken from a $N(r,s)$ distribution such as the $N(5,1)$ shown on the left in graph 1, but under H_1, the Y sample comes from the $N(r+0.61s, s)$ distribution, while the X sample comes from the $N(r,s)$ distribution.

Another example of a possible situation satisfying the alternative hypothesis, H_1, given approximately by comparing a gamma (4,1) with a gamma (3,1), is illustrated by graph 2.

Note that the Wilcoxon Rank Sum Test is most sensitive to location, a little sensitive to shape, but not to dispersion (except as it relates proportionately to differences in location). Therefore, it is the differences in location that are of primary importance in graphs 1 and 2.

In order to determine the probability of drawing a value from distribution A which is larger than a simultaneously drawn value from distribution B, the following may be used:

$$P = \int_{x=-\infty}^{\infty} f_B(x) \int_{t=x}^{\infty} f_A(t)\,dt\,dx$$

where f_A and f_B represent density functions.

For the case where A and B are both gamma distributions,

$$P = 1 - \beta_A^{-\alpha_A} \beta_B^{-\alpha_B} \frac{1}{(\alpha_A - 1)!} \sum_{r=1}^{\alpha_B - 1} \frac{\beta_B^{r+1}}{(\alpha_B - 1 - r)!} \frac{(\alpha_A + \alpha_B - 2 - r)!}{([\alpha_A + \beta_B]/\beta_A\beta_B)^{\alpha_A + \alpha_B - 1 - r}}$$

For gamma (4,1) and gamma (3,1), $P = 21/32 \approx 0.656$.

For normal distributions, use $\Phi[(\mu_A - \mu_B)/\sqrt{\sigma_A^2 + \sigma_B^2}]$, as in the Church-Harris-Downton (C-H-D) method of missile motor safety testing [2]. (Note: This reference to the C-H-D method should not be construed as the author's endorsement of this method for the purpose of missile motor safety testing.)

The calculation of power under this alternative involves a summation over a typically large number of products. Calculation of this value can become extremely time consuming, even for a high speed computer. A program was written for the author at White Sands Missile Range which will calculate these exact values, however, in general, the sample sizes must be very small. Recently, however, the author constructed a simulation which provides estimates of the power for much larger sample sizes. A number of the "products" mentioned earlier are calculated and the mean is computed. The number of products involved in the exact calculation can be determined, and it is multiplied by this mean. Comparison to values calculated exactly (when practical), and a study of the sensitivity of the results to increased replications, as well as comparison to other simulated values bounding the results in the tables, led to the use of from 1 to 20 million replications to simulate values for the tables found in this paper. (Work has been done, reference [3], to determine the number of simulation replications needed under less radical circumstances. Here, however, a larger number of replications appears necessary.) (For n = m

= 50, up to 35 million replications were used. It appeared, however, that fewer replications using a number of different seeds yielded mean answers which more quickly converged to reasonable results, especially when using antithetic seeds.)

In the tables, n is the sample size of the X sample, m is the sample size of the Y sample, RS is the rank sum for which type I and type II error probabilities are calculated, PA is the former of those probabilities, and PB is the later. Specifically, PA is the attained probability of making an error if H_0 is rejected, and PB is the attained probability of error if H_1 is rejected, both corresponding to the same RS value. RS is always calculated by adding the ranks of the Y elements in the combined sample. Note that for smaller sample sizes, power +PB is noticeably larger than unity due to the discrete nature of this test. That is, the probability of obtaining exactly the event observed (and no other) is non-zero.

Three significant digits are given for PA and only two for power and PB simply because it takes fewer replications of the simulation to satisfactorily obtain a value for PA than for the others.

From the annex to table 1, it is found empirically that if x is the size of each of the two samples, and $f_\alpha(x)$ is the probability of a type II error under the alternative used here, adjusted to correspond to a specific significance level, then, as a continuous representation of actually a discrete process,

$$f_{0.10}(x) \approx \exp(-x/16)$$

for at least $3 \le x \le 40$, and perhaps this approximation could be trusted for x = 45 or larger. However, extrapolations are always more dangerous than interpolations, so caution is advised for further extensions.

For $\alpha = 0.05$,

$$f_{0.05}(x) \approx \exp(-x/[26\exp \frac{10-x}{5x}])$$

for at least $4 \le x \le 40$, and perhaps for x substantially larger. Using this approximation, it is conjectured that for n = m = 66, when PA is approximately 0.05 (RS = 4751), then PB for this alternative is also approximately 0.05 and the true state of nature would then quite safely be said to (probably) lie between the null and alternative hypotheses. (At the 0.1 probability level for PA and PB, this could be said when n = m = 37, and RS = 1507.) An extrapolation to n = m = 66 is questionable, however, and further extrapolation is not advised. Computer simulation for n = m = 50 indicates that for the top curve (PA $\simeq$ 0.05) in Annex I to table 1, true values in this area for PB may be somewhat smaller than this curve predicts. For PA $\simeq$ 0.10, PB values for large n and m may be somewhat larger than predicted.

In Conover's book [1], an approximation is given to find RS for a given PA value. ($RS \gtrsim m(m + n + 1)/2 + x_{1-\alpha}\sqrt{mn(n + m + 1)/12}$, where $x_{1-\alpha}$ is from the table of the cumulative normal distribution.) The two functions given earlier can be used to estimate PB values when PA $\approx$ 0.10 or 0.05.

The final graphs, 3-7, are taken from work the author directed at White Sands Missile Range in order to study this alternative for the Wilcoxon Rank Sum Test with emphasis on simulation validation for missile flight simulations. When comparing a very few live firings to a substantially larger number of simulations for each scenario, it can be seen from these graphs that once one sample is substantially larger than the other, increasing the larger sample size further does very little to improve the power. These graphs are continuous representations of what are actually discrete points. The values for those points were calculated analytically as noted in the acknowledgements.

Finally, when $n \neq m$, PB can be bounded using the exponential formulations found earlier in this paper. If, for example, RS is such that PA = 0.1, and x_1, is the smaller of n and m, and x_2 is the larger, then one has that approximately $\exp(-x_2/16) < PB < \exp(-x_1)$, with PB somewhat closer to $\exp(-x_1/16)$, especially when $x_1 \ll x_2$.

For larger sample sizes than are handled here, parametric methods may be used. However, in addition to the probability of error associated with any conclusion drawn from a parametric test, there is the additional risk involved in assuming the distributional forms used in such a test. Hypothesis tests should also be used to study these distributional assumptions to provide a more complete risk analysis.

EXAMPLE:

Consider two sources of data, X and Y, where it is suspected that Y may represent a population of larger location than X, but this is not clear. If 11 observations are taken from the X population, and 19 observations taken from Y, then the critical value of the rank sum (RS) of the Y sample observations within the combined sample which represents the point at which rejection of the null hypothesis would occur using α = 0.10, is approximately

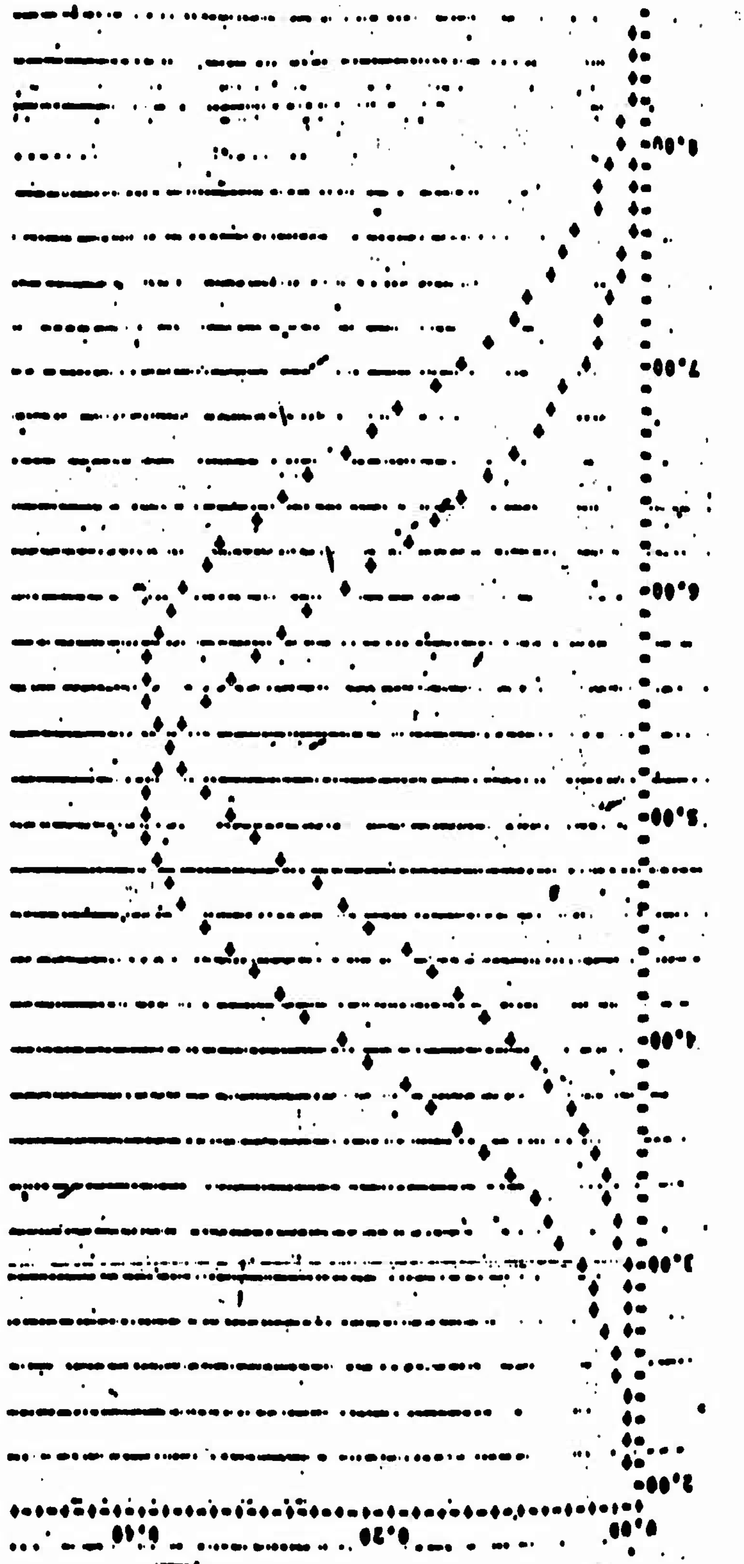
$$RS \gtrsim m(m + n + 1)/2 + 1.2816\sqrt{mn(m + n + 1)/12}$$

$$= (19)(31)/2 + 1.2816\sqrt{(19)(11)(31)/12}$$

$$\approx 324.3$$

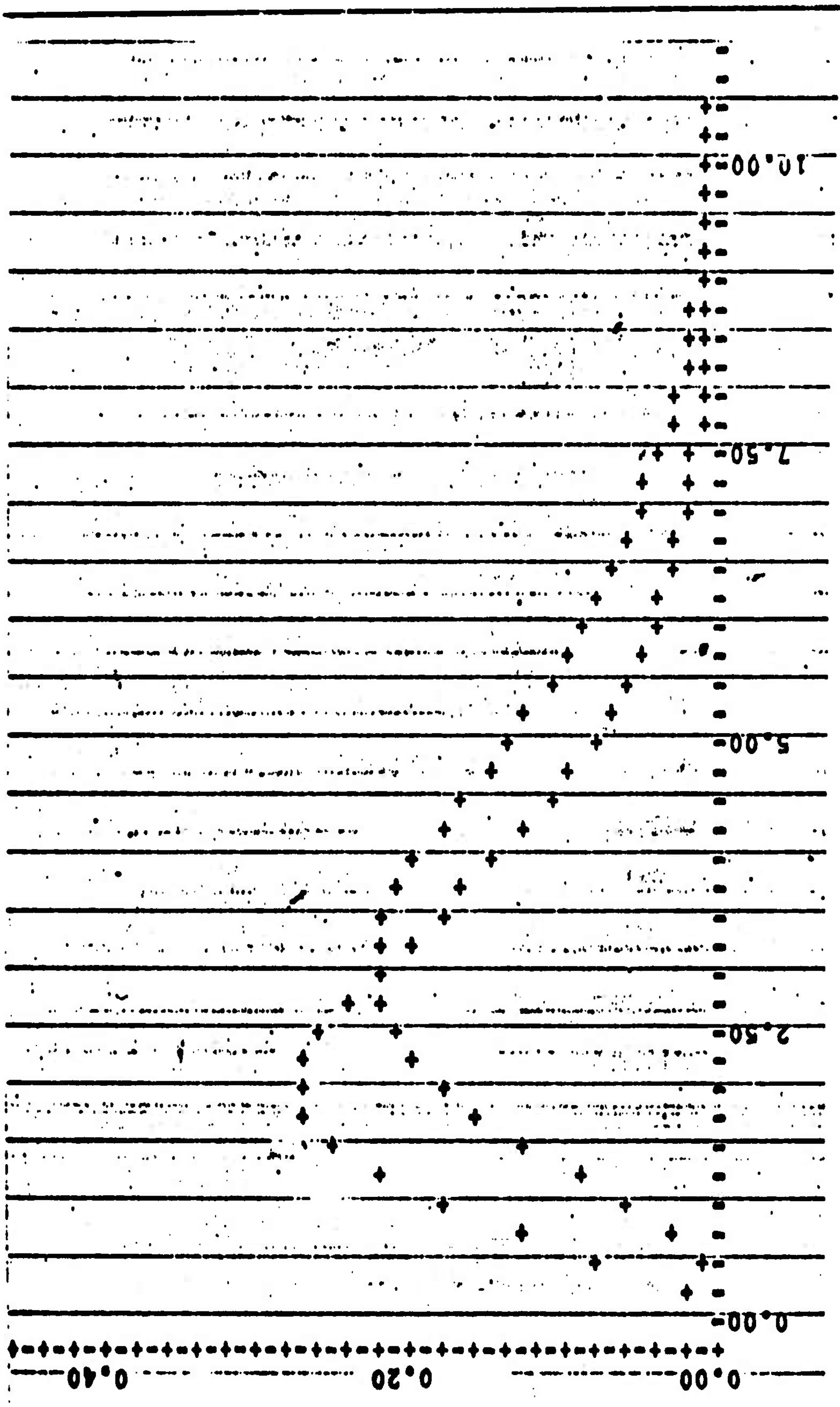
Therefore, if $RS \geq 325$, H_0 would be rejected at the α = 0.10 level. However, should RS = 325, and H_0 not be rejected, then the probability of making a type II error with respect to the alternative hypothesis illustrated in graphs 1 and

2 is approximately bounded by exp (-19/16) and exp (-11/16), so 0.30<PB<0.50. Note that, from table 2 , when PA = 0.099, PB (10,20) ≈ 0.43. Using 4,000,000 replications in the program given in Appendix A, for m = 19, n = 11, and RS = 325, resulted in PA = 0.100 and PB = 0.42.



Alternative Hypothesis (H_1) Using a
N(5,1) and an N(5.61,1)

Graph 1

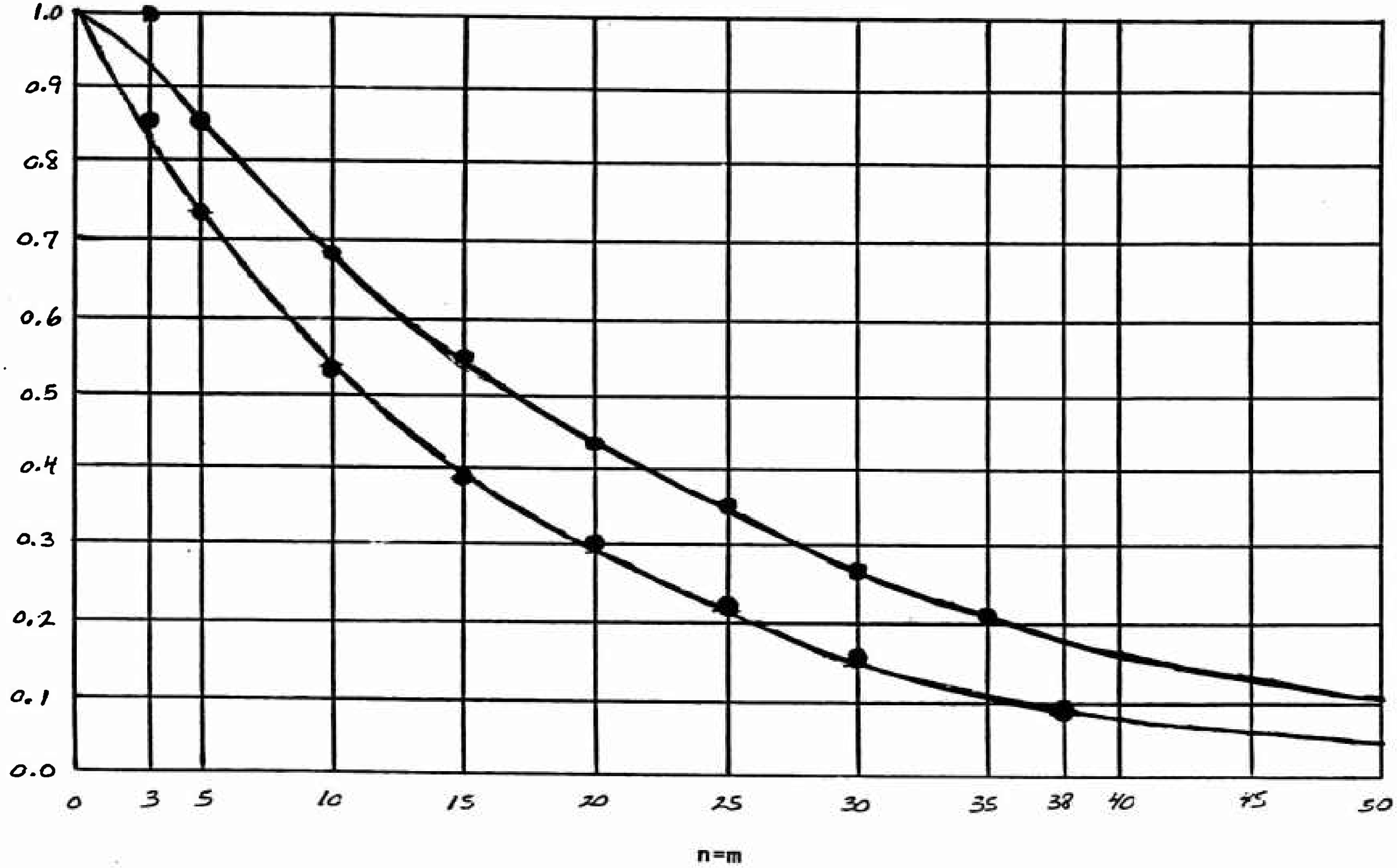


Graph 2

Table 1

n = m	RS	PA	power	PB
3	12	0.350	0.62	0.56
3	14	0.100	0.27	0.85
3	15	0.050	0.15	1.00
5	32	0.210	0.54	0.55
5	34	0.111	0.37	0.71
5	35	0.075	0.29	0.79
5	36	0.048	0.21	0.86
5	39	0.008	0.05	0.97
10	122	0.108	0.52	0.51
10	123	0.095	0.49	0.54
10	127	0.052	0.36	0.67
10	128	0.045	0.34	0.69
10	136	0.009	0.13	0.89
15	264	0.101	0.63	0.39
15	265	0.094	0.61	0.41
15	273	0.049	0.47	0.55
15	289	0.009	0.21	0.80
20	458	0.101	0.71	0.30
20	459	0.096	0.70	0.30
20	471	0.051	0.58	0.43
20	472	0.048	0.57	0.44
20	496	0.010	0.30	0.71
25	704	0.101	0.79	0.22
25	705	0.098	0.78	0.23
25	723	0.050	0.66	0.35
25	758	0.009	0.38	0.63
30	1002	0.101	0.85	0.16
30	1003	0.099	0.84	0.16
30	1027	0.050	0.74	0.27
30	1073	0.010	0.47	0.54
35	1383	0.050	0.79	0.21
38	1587	0.100	0.91	0.09

PB FOR PA≈0.05 (TOP) AND PA≈0.10 (BOTTOM)



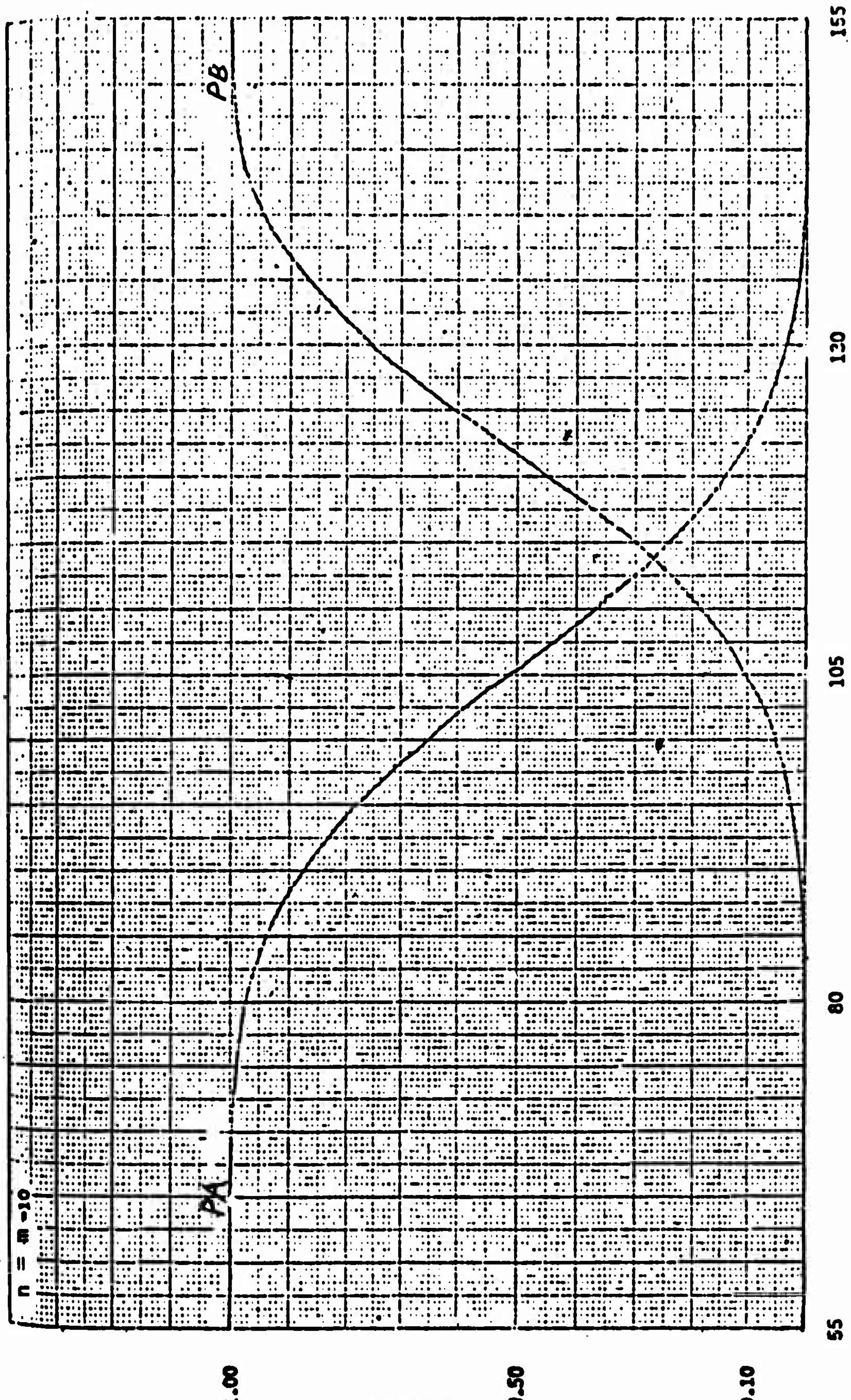
POINTS GRAPHED ARE FROM SIMULATION

TOP CURVE IS $f_{0.05}(x) = \exp[-x/(26\exp\frac{10-x}{5x})]$

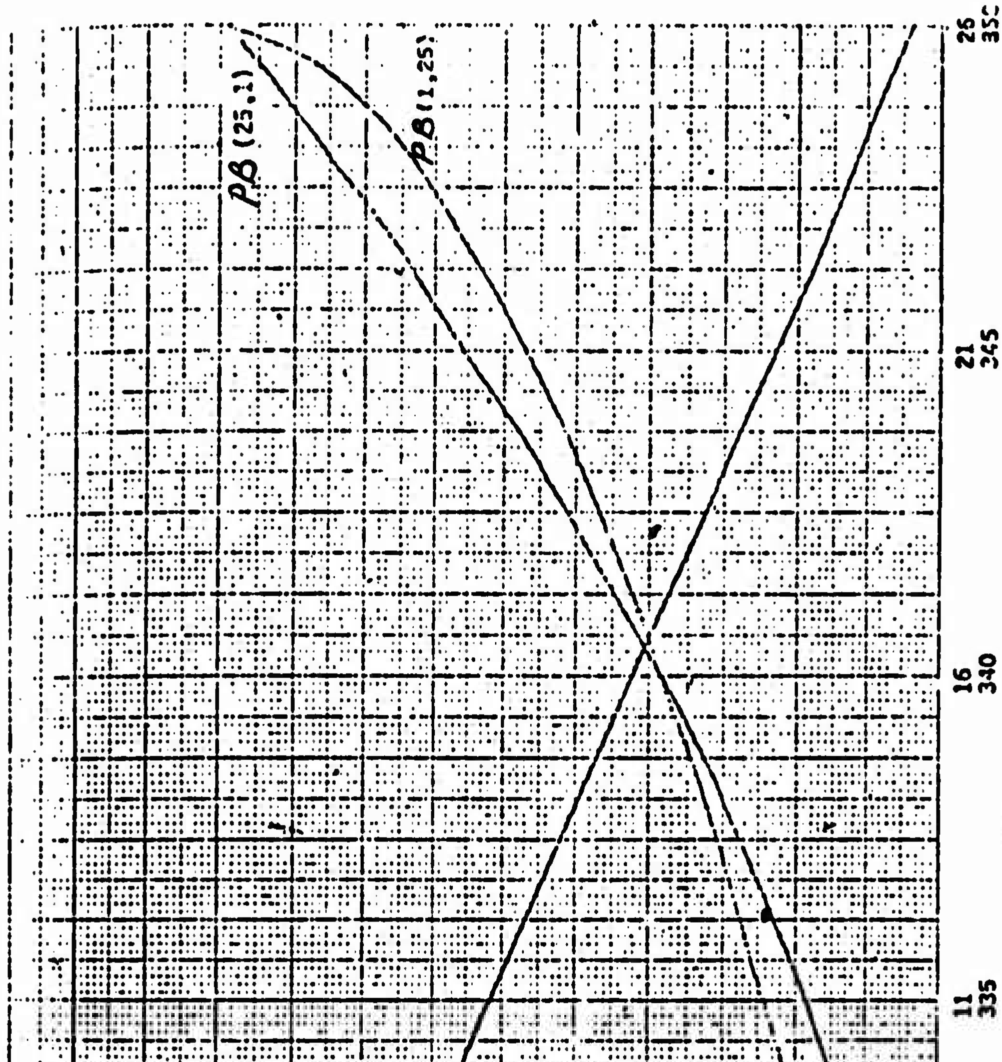
BOTTOM CURVE IS $f_{0.10}(x) = \exp(-x/16)$

Table 2

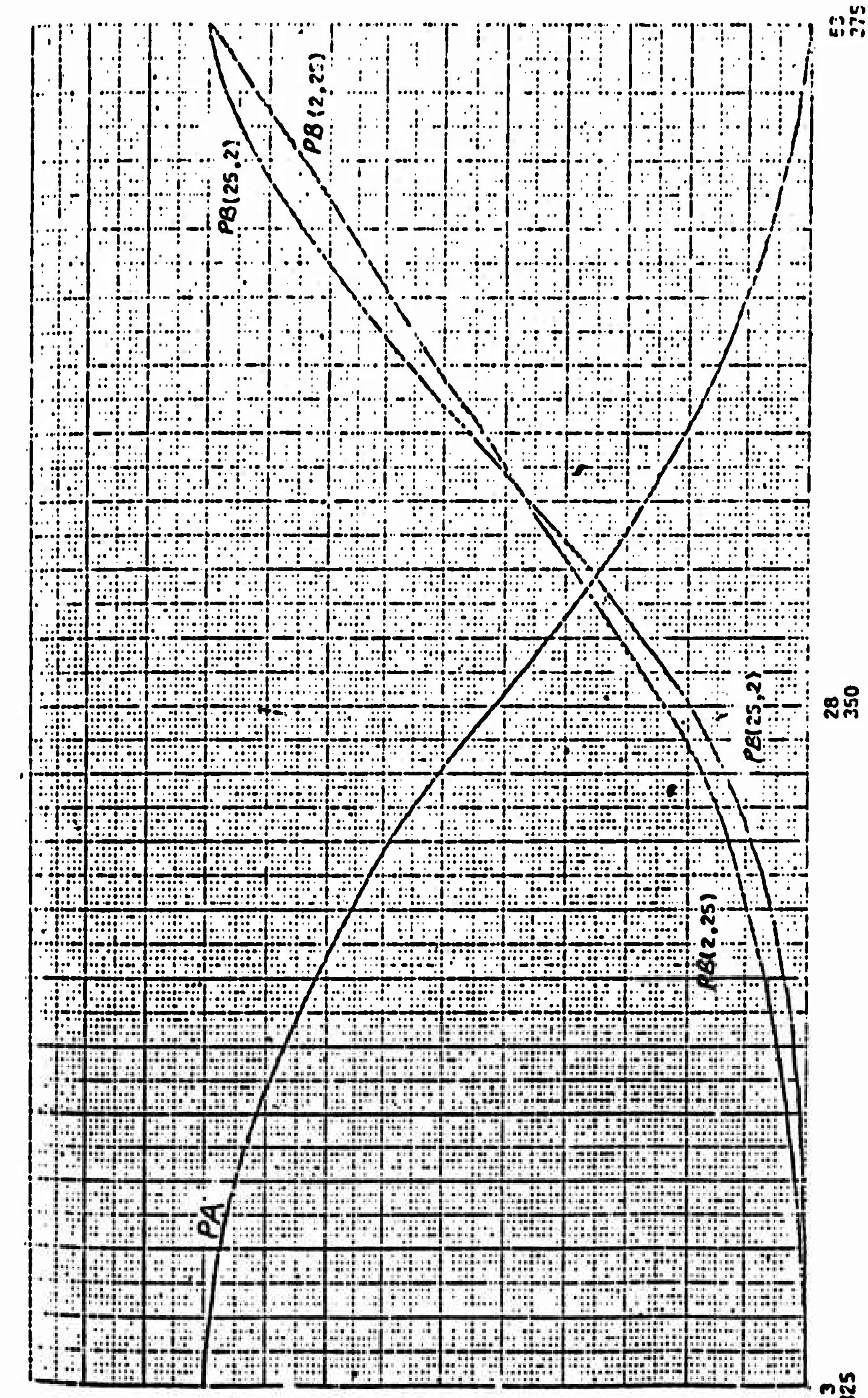
n, m n ≠ m	RS	PA	power	PB
5,10	85	0.297	0.70	0.35
5,10	91	0.103	0.42	0.63
5,10	92	0.082	0.37	0.68
5,10	94	0.050	0.27	0.77
5,10	99	0.010	0.10	0.93
10,5	45	0.297	0.71	0.34
10,5	51	0.103	0.41	0.65
10,5	52	0.082	0.35	0.70
10,5	54	0.050	0.26	0.79
10,5	59	0.010	0.08	0.95
5,25	412	0.094	0.45	0.57
5,25	418	0.048	0.33	0.69
5,25	429	0.009	0.13	0.88
5,25	430	0.008	0.12	0.89
25,5	102	0.094	0.44	0.59
25,5	108	0.048	0.29	0.73
25,5	119	0.009	0.09	0.92
25,5	120	0.008	0.08	0.93
10,20	340	0.099	0.58	0.43
10,20	348	0.050	0.44	0.58
10,20	363	0.009	0.20	0.81
20,10	185	0.099	0.59	0.43
20,10	193	0.050	0.44	0.58
20,10	208	0.010	0.18	0.84
5,50	1444	0.105	0.50	0.51
5,50	1457	0.050	0.36	0.65
5,50	1480	0.008	0.14	0.87
50,5	184	0.105	0.50	0.52
50,5	197	0.050	0.32	0.69
50,5	220	0.008	0.09	0.92
10,50	1590	0.101	0.65	0.36
10,50	1608	0.051	0.52	0.49
10,50	1643	0.009	0.26	0.75
50,10	370	0.102	0.68	0.33
50,10	388	0.051	0.52	0.49
50,10	423	0.009	0.22	0.79



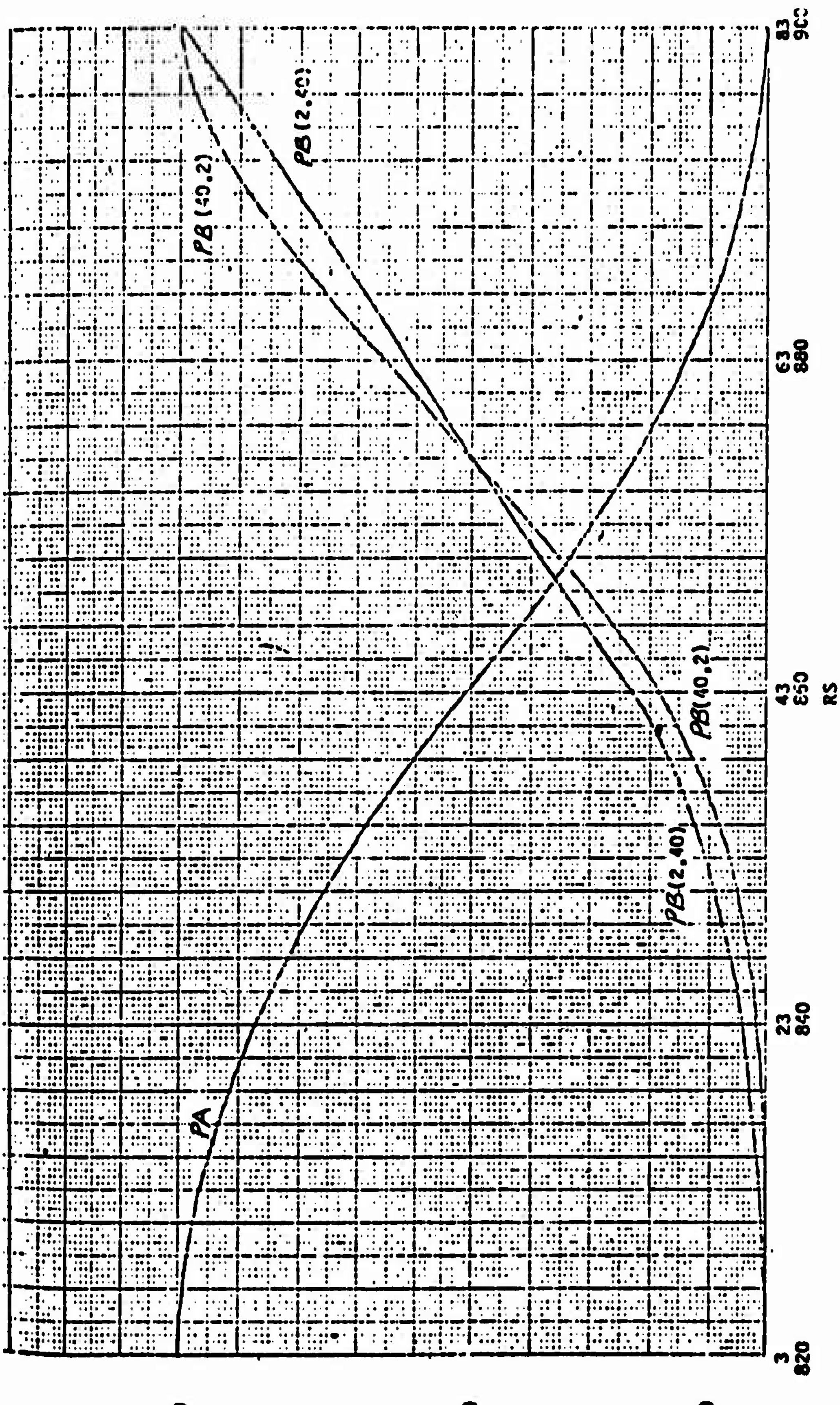
Graph 3


1.00
0.50
0.10
RS
1
325
6
330
11
335
16
340
21
345
26
350
PA
PB(25,1)
PB(1,25)

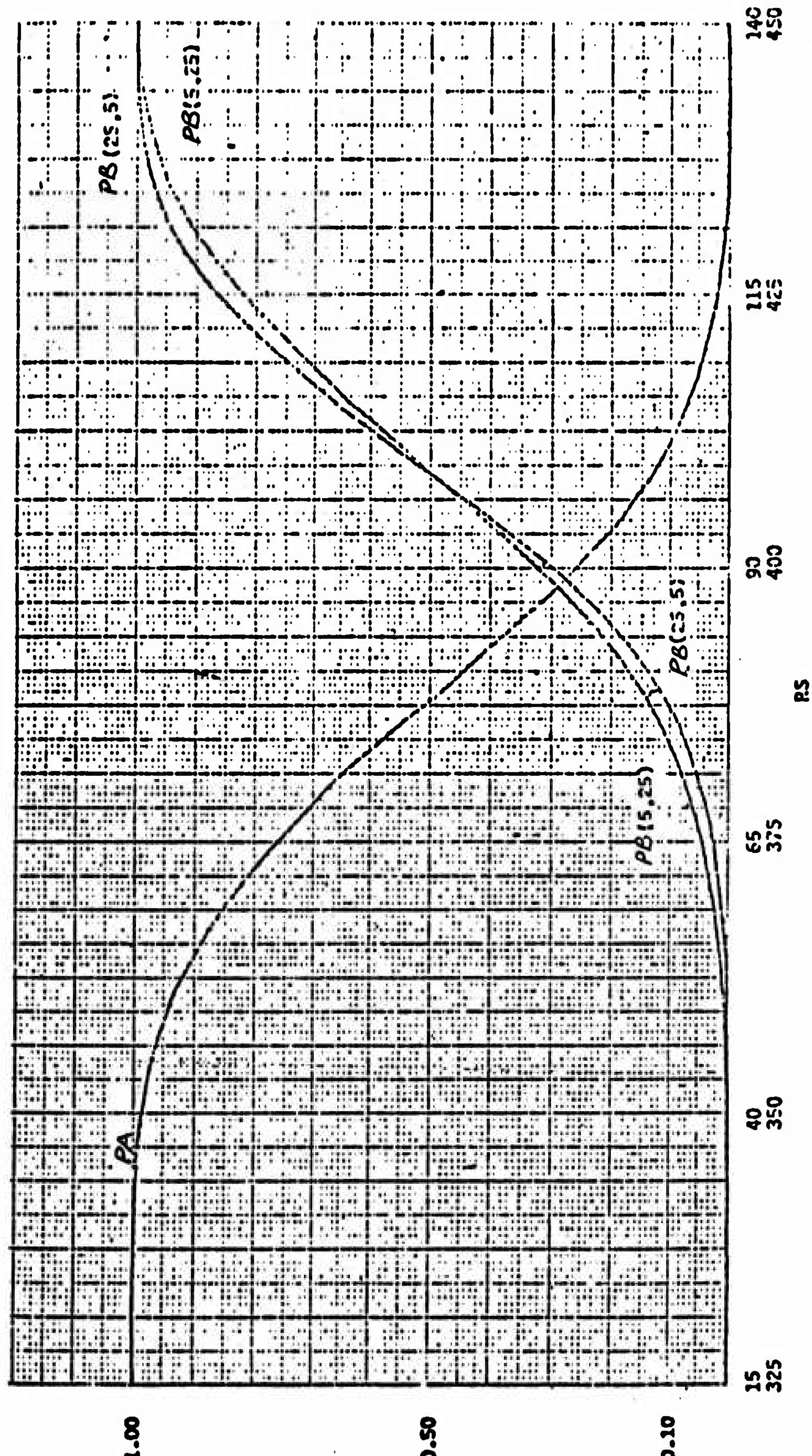
Graph 4



Graph 5


PB(40,2)
PB(2,40)
PB(2,40)
PB(40,2)
PA
1.00
0.50
0.10
3
820
23
840
43
860
63
880
83
900
RS

Graph 6



PA
PB(25,5)
PB(5,25)
RS
1.00
0.50
0.10
15
325
40
350
65
375
90
400
115
425
140
450

Graph 7

APPENDIX A

FORTRAN CODE FOR

SIMULATION:

"LEHMANN POWER ANALYSIS

FOR THE

WILCOXON RANK SUM TEST"

(LPAWRST)

VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST
VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST
VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST

```
K   K  N   N   AAA   U   U  BBBB
K  K   N   N  A   A  U   U  B   B
K K    NN  N  A   A  U   U  B   B
KKK    N N N  A   A  U   U  BBBB
K K    N  NN  AAAAA  U   U  B   B
K  K   N   N  A   A  U   U  B   B
K   K  N   N  A   A  UUUUU  BBBB
```

```
LL          PPPPPPPP      AAAAAA    WW      WW  RRRRRRRR      SSSSSSSS  TTTTTTTTTT
LL          PPPPPPPP      AAAAAA    WW      WW  RRRRRRRR      SSSSSSSS  TTTTTTTTTT
LL          PP      PP  AA      AA  WW      WW  RR      RR  SS              TT
LL          PP      PP  AA      AA  WW      WW  RR      RR  SS              TT
LL          PP      PP  AA      AA  WW      WW  RR      RR  SS              TT
LL          PP      PP  AA      AA  WW      WW  RR      RR  SS              TT
LL          PPPPPPPP    AA      AA  WW      WW  RRRRRRRR      SSSSSS        TT
LL          PPPPPPPP    AA      AA  WW      WW  RRRRRRRR      SSSSSS        TT
LL          PP          AAAAAAAAAA  WW  WW  WW  RR  RR              SS      TT
LL          PP          AAAAAAAAAA  WW  WW  WW  RR  RR              SS      TT
LL          PP          AA      AA  WWWW  WWWW  RR    RR            SS      TT      ...
LL          PP          AA      AA  WWWW  WWWW  RR    RR            SS      TT      ...
LLLLLLLLLL  PP          AA      AA  WW      WW  RR      RR  SSSSSSSS        TT      ...
LLLLLLLLLL  PP          AA      AA  WW      WW  RR      RR  SSSSSSSS        TT      ...
```

```
FFFFFFFFFF    OOOOOO    RRRRRRRR        ;;;;      222222
FFFFFFFFFF    OOOOOO    RRRRRRRR        ;;;;      222222
FF          OO      OO  RR      RR      ;;;;    22      22
FF          OO      OO  RR      RR      ;;;;    22      22
FF          OO      OO  RR      RR                      22
FF          OO      OO  RR      RR                      22
FFFFFFFF    OO      OO  RRRRRRRR        ;;;;          22
FFFFFFFF    OO      OO  RRRRRRRR        ;;;;          22
FF          OO      OO  RR  RR          ;;;;        22
FF          OO      OO  RR  RR          ;;;;        22
FF          OO      OO  RR    RR          ;;      22
FF          OO      OO  RR    RR          ;;      22
FF            OOOOOO    RR      RR      ;;      2222222222
FF            OOOOOO    RR      RR      ;;      2222222222
```

```
K   K  N   N   AAA   U   U  BBBB
K  K   N   N  A   A  U   U  B   B
K K    NN  N  A   A  U   U  B   B
KKK    N N N  A   A  U   U  BBBB
K K    N  NN  AAAAA  U   U  B   B
K  K   N   N  A   A  U   U  B   B
K   K  N   N  A   A  UUUUU  BBBB
```

VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST
VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST
VAX/VMS KNAUB LPA*RST 14-JUL-1983 15:35 LPA0: 14-JUL-1983 15:35 USER:[KNAUB]LPA*RST

```
      INTEGER PREV(317)
      DIMENSION I(1000),NEXT(1000),II(3),MIN(317),MAX(317)
      LOGICAL*1 FLAG(317)
      REAL*4 I,II,MINVAL,MAXVAL
      REAL*8 BOUND(318),XX,PS,TPS,PTPS,PBC,PBP,C2,PBC1,PBC2,
     /   PBC3,PBC4
      DATA ISEED/78125/,IDIR/1/
      M=0.
      WRITE(19,111)
111   FORMAT(1X,'LEHMANN POWER ANALYSIS FOR THE WILCOXON
     /RANK SUM TEST, LPAWRST')
      WRITE(19,1)
      WRITE(6,1)
1     FORMAT(1X,'ENTER NO. OF OBSERVATIONS, NO. Y')
      READ(5,*)NOBS,NY
      WRITE(19,*)NOBS,NY
      WRITE(19,101)
   7   PRINT 101
101    FORMAT(1X,'INPUT TEST STATISTIC')
       READ(5,*)IX
       WRITE(19,*)IX
       WRITE(19,102)
      PRINT 102
102   FORMAT(1X,'INPUT NO. OF REPLICATIONS')
      READ(5,*)IREPS
      WRITE(19,*)IREPS
100   I(1)=RAN(ISEED)
      IYRNK=0
      DO 105 J=1,317
      FLAG(J)=.FALSE.
105   CONTINUE
      MINVAL=I(1)*IDIR
      MAXVAL=0
      DO 10 J=2,NOBS
      I(J)=RAN(ISEED)
      II(1)=I(J)*IDIR
      IF(II(1).GT.MAXVAL)MAXVAL=II(1)
      IF(II(1).LT.MINVAL)MINVAL=II(1)
10    CONTINUE
      J=J-1
      NUM=J
      X=NUM
      NCELLS=(X/(SQRT(X)))+.5
      RANGE=MAXVAL-MINVAL
      BOUND(1)=MINVAL
      XX=RANGE/NCELLS
      DO 60 J=2,NCELLS
      BOUND(J)=BOUND(J-1)+XX
60    CONTINUE
      BOUND(NCELLS+1)=MAXVAL
      DO 50 J=1,NUM
      XX=I(J)*IDIR
      DO 70 JJ=1,NCELLS
      IF(XX.GE.BOUND(JJ).AND.XX.LT.BOUND(JJ+1))GO TO 75
70    CONTINUE
      JJ=JJ-1
75    IF(FLAG(JJ).EQ..FALSE.)THEN
         MIN(JJ)=J
         MAX(JJ)=J
         FLAG(JJ)=.TRUE.
         GO TO 50
```

```
      END IF
      II(1)=XX
      II(2)=I(MIN(JJ))*IDIR
      II(3)=I(MAX(JJ))*IDIR
      IF(II(1).LE.II(2))THEN
            NEXT(J)=MIN(JJ)
         MIN(JJ)=J
      ELSE IF(II(1).GT.II(3))THEN
         NEXT(MAX(JJ))=J
         MAX(JJ)=J
      ELSE
         PREV(JJ)=MIN(JJ)
         K=NEXT(MIN(JJ))
20       II(1)=I(J)*IDIR
         II(2)=I(K)*IDIR
         IF(II(1).LE.II(2))GO TO 30
         PREV(JJ)=K
         K=NEXT(K)
         GO TO 20
30       NEXT(PREV(JJ))=J
         NEXT(J)=K
      END IF
50    CONTINUE
      L=0
      PS=1.
      IIY=0
      DO 80 JJ=1,NCELLS
      IF(FLAG(JJ).EQ..FALSE.)GO TO 80
      K=MIN(JJ)
40    IIY=IIY+1
      IF(K.LE.NY)THEN
         L=L+1
         IYRNK=IYRNK+IIY
         PS=PS*(IIY+L-1)/100.
         IALPH='Y'
      ELSE
         IALPH='X'
      END IF
C     WRITE(6,2)I(K),IALPH
2     FORMAT(1X,E15.7,5X,A1)
      IF(K.EQ.MAX(JJ))GO TO 80
      K=NEXT(K)
      GO TO 40
80    CONTINUE
C      WRITE(6,3)IYRNK
C3     FORMAT(1X,'SUM OF Y - RANKS = ',I6)
       IF(IYRNK.GE.IX)M=M+1
        IF(IYRNK.EQ.IX)THEN
                    PTPS=PTPS+PS
                    NPTPS=NPTPS+1
              END IF
       IF(IYRNK.LE.IX)GO TO 200
       TPS=TPS+PS
       NPTPS=NPTPS+1
       NTPS=NTPS+1
       PTPS=PTPS+PS
200    ITRACK=ITRACK+1
       XM=M
       IF(ITRACK-IREPS)100,201,201
201    XNOBS=NOBS
       XNY=NY
```

```
      C1=TREPS
      PA=XM/C1
      XNTPS=NTPS
       WRITE(19,*)XNTPS
      XNPTPS=NPTPS
        IF(TPS.EQ.0.)THEN
             ATPS=0.
        'ELSE
         ATPS=TPS/XNTPS
               END IF
        IF(PTPS.EQ.0.)THEN
             APTPS=0.
         ELSE
         APTPS=PTPS/XNPTPS
               END IF
      C2=(XNOBS+XNY)*(10.**20)
      DO 203 IL=1,NY-1
      XI=IL
      C2=C2*(XNOBS+XNY-XI)/100.
203   CONTINUE
C        WRITE(6,*)C2
      XNYF=1.
        PB=ATPS*PA
         PBP=APTPS*PA
        PBC1=(2.**30)*(10.**20)
        PBC2=(2.**30)*PBP
        PBC3=(10.**2)*(2.**(NY-60))
        PBC4=PBC3/C2
         POWER=PBC1*PBC2*PBC4
       WRITE(19,*)ITY,L,PB,C2
       IF(ATPS.EQ.0.)THEN
                 PB=0.
             ELSE
         PB=POWER*(PB/PBP)*(XNTPS/XM)
           END IF
900      PB=1.0-PB
      WRITE(6,*)PA,POWER,PB
      WRITE(19,*)PA,POWER,PB
      STOP
      END
```

APPENDIX B

ACKNOWLEDGEMENTS

Thanks to Lynne Grile and Gerard Petet for developing the programming necessary to generate graphs 1 and 2. Keith Haycock (White Sands) constructed graphs 3-7, and Dr. Larry Armijo (White Sands - KENTRON) wrote the computer program to provide analytical solutions used in graphs 3-7. Also, thanks to Jeffrey Greenhill for providing a customized sorting routine for the author's simulation. Carlyle Comer and Frank Lawrence were helpful in obtaining massive quantities of needed CPU time on the USALOGC's VAX-11/780. Finally, thanks to other analysts for helpful conversations and/or influence.

APPENDIX C

REFERENCES

1. Conover, W. J., Practical Nonparametric Statistics, 2 ed, John Wiley & Sons, 1980.

2. Downs, R. S., P. C. Cox, "The Probability of Motor Case Rupture," ARO Report 75-2.

3. Juritz, J. M., J. W. F. Juritz and M. A. Stephens, "On the Accuracy of Simulated Percentage Points," Journal of the American Statistical Association, 78 (June 1983).

4. Knaub, J. R., Jr., "Design of a Multiple Sample Westenberg Type Test for Small Sample Sizes," ARO Report 82-2.

5. Knaub, J. R., Jr., Appendix D, US Army Manpower Nonavailability and Indirect (Unit Related) Productive Factors - Final Report, US Army Logistics Center, August 1983 - not released

6. Knaub, J. R., Jr., L. M. Grile and G. Petet, "Analyzing n Samples of 2 Observations Each," ARO Report 83-2.

7. Lehmann, E. L., "The Power of Rank Tests," Annals of Mathematical Statistics, 24 (1953), 23-43.

ADDENDUM

Multiple applications of this test can be used to compare two levels of a factor under a number of conditions. If, for example, manufacturer A produces a machine which is suspected to have higher reliability under most scenarios than a similar machine made by manufacturer B, then under each of the γ scenarios, m_i is the sample size of A's machines and n_i is the sample size of B's machines, for $i = 1$ to γ. PA_i and PB_i can be calculated for each of the scenarios. Consider $0 \leq a \leq \gamma$ and $0 \leq b \leq \gamma$.

PA is the probability of a or more PA_i's being less than p_A ($i = 1, \gamma$), when H_0 is true.

PB is the probability of b or more PB_i's being less than p_B ($i = 1, \gamma$), when H_1 is true.

Therefore,

$$PA = \sum_{x=a}^{\gamma} \binom{\gamma}{x} P_A^x (1 - P_A)^{\gamma - x}$$

and

$$PB = \sum_{x=b}^{\gamma} \binom{\gamma}{x} P_B^x (1 - P_B)^{\gamma - x}$$

P_A and P_B are chosen to be reasonable considering sample sizes for each of the γ cases.

If $\frac{PA}{PB} = 1$ then the evidence shows that, in general, the true state of nature is just as likely to be equivalent to H_1 as H_0.

If $\frac{PA}{PB} = 2$ then the evidence indicates that, in general, the true state of nature is twice as likely to be equivalent to H_0 as H_1. If *PA* and *PB* are small, then the indication is only that the true state of nature is closer to H_0 than H_1, although possibly not very close to either.

(Note that another paper in this conference, "Numerical Validation of Tukey's Criteria for Clinical Trials and Sequential Testing," by C. R. Leake, also deals with this type of problem, and was of interest to this author.)

At this time, this methodology is being used to determine whether survey data from a presumably less reliable source is compatible with a presumably superior data source. Difficult to obtain data on U.S. Army warehousing activities have, as one obvious characteristic, a very flat "peak." Therefore, a sample median value can be changed drastically by the addition or deletion of one data point. If the secondary data source proves to provide values distributed closely enough to that of the primary source, the advantage of including this source may outweigh the disadvantage. The current situation is more complex than this. However, some results employing the methodology of this addendum have been realized.

ADDENDUM 2

Two approximations for the power of this test which apparently are good for a wide range of normal alternative hypotheses are to be found in E. L. Lehmann, Nonparametrics: Statistical Methods Based on Ranks, Holden-Day, 1975. Although restricted to normal alternatives in the format in which they are written, these approximations can be used to extend the tables given here to larger n and m. The easier of the two approximations to apply, in its simplest form, is found on page 73 of the above reference and is essentially as follows:

$$\text{power} \approx \Phi\left[\sqrt{\frac{3mn}{(m+n+1)\pi}}\,\frac{\mu_A - \mu_B}{\sigma} - \chi_{1-\alpha}\right]$$

where in our case we have $(\mu_A - \mu_B)/\sigma \approx 0.610$.

Note that in the example in the main body of this paper ($m = 19, n = 11$), that this approximation gives power ≈ 0.60, which is consistent with what was shown earlier.