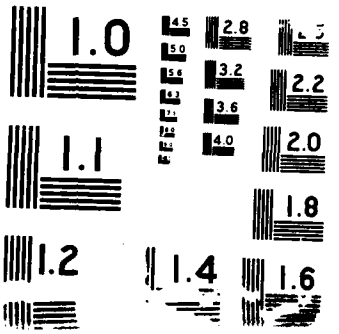
AD-A194 308 A MINITAB MACRO FOR THE KAPLAN-MEIER PRODUCT LIMIT ESTIMATOR AND GEHAN'S TEST(U) WISCONSIN UNIV-MADISON CENTER FOR MATHEMATICAL SCIENCES A P SOMS NOV 87 1/1

UNCLASSIFIED CMS-TSR-88-15 F/G 12/5 NL

1.0
1.1
1.2
1.4
1.6
1.8
2.0
2.2
2.8
3.2
3.6
4.0
4.5
5.0
5.6

DTIC FILE COPY

AD-A194 300

CMS Technical Summary Report #88-15

# A MINITAB MACRO FOR THE KAPLAN-MEIER PRODUCT LIMIT ESTIMATOR AND GEHAN'S TEST

Andrew P. Soms

UNIVERSITY OF WISCONSIN

CENTER FOR THE MATHEMATICAL SCIENCES

**Center for the Mathematical Sciences**
**University of Wisconsin—Madison**
**610 Walnut Street**
**Madison, Wisconsin 53705**

November 1987

(Received November 2, 1987)

**Approved for public release**
**Distribution unlimited**

DTIC
ELECTE
APR 19 1988
S
E

88 3 23 046

UNIVERSITY OF WISCONSIN-MADISON
CENTER FOR THE MATHEMATICAL SCIENCES

# A MINITAB MACRO FOR THE KAPLAN-MEIER PRODUCT LIMIT ESTIMATOR AND GEHAN'S TEST

Andrew P. Soms*

Technical Summary Report #88-15
November 1987

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Code
Dist | Avail and/or Special
A-1

## ABSTRACT

Two versions of a Minitab macro to compute the K-M ple and Gehan's test are given. The shorter, km.geh, prints the K-M ple for both groups, gives simultaneous plots of the survival curves and calculates the one and two-sided P level of Gehan's test. The longer, kmt.geh, in addition to the above, prints an ordered table of the observed values and their Gehan scores.

DTIC COPY INSPECTED 4

AMS (MOS) Subject Classification: 62-04

Key Words: Gehan's test; Kaplan-Meier; Macro; Minitab; Product-limit estimator

*Department of Mathematics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

# A MINITAB MACRO FOR THE KAPLAN-MEIER PRODUCT LIMIT ESTIMATOR AND GEHAN'S TEST

Andrew P. Soms*

## 1. Introduction

Minitab is a user-friendly statistical computer package widely used in industry and academia. For more details, see Ryan, Joiner and Ryan (1985) or Cobb and Gifford (1986). The Kaplan-Meier ple and Gehan's test are often used in biostatistics and reliability theory to compare two survival distributions. For definitions, assumptions and examples, see Brown, and Hollander (1977) or Miller (1981).

The main idea of the program is to first break ties between censored and completed observations by adding to the censored observations one-half of the smallest non-zero difference for the ordered combined observations. This does not change the K-M ple or Gehan's test. Then, using an extension of "Macro for obtaining distinct values and replicate counts from a column of replicated data" by Gunter, Minitab Users' Group Newsletter, 5 (1984), 1, an ordered table of distinct values, together with total replications, type of observation (censored or completed), number in group one and number in group two, is constructed. From this, the K-M ple and Gehan's test are then readily computed. The second version of the program, kmt.geh, in addition to doing everything the first, km.geh, does, also prints out the table and Gehan's scores.

---

*Department of Mathematics, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.

An example from Miller (1981, pp. 49-50, pp. 203-205) has been worked in the Appendix and agrees with Miller's solution. We also checked the example in Brown and Hollander (1977), p. 345 and pp. 347-350, and found agreement.

The two macros listed in the Appendix can be modified by the user, if desired, by his computer's editor. The data can be conveniently read in from a file by the Minitab 'read' command, as was done for the example in the Appendix. The macros can easily be used by scientists not familiar with statistics and/or more complex statistical computer packages to quickly assess the significance of their data. For University of Wisconsin-Madison users, it may be possible to transfer the macros directly from the Center for the Mathematical Sciences VAX 780 to the user's computer and interested persons should contact the author.

The author would appreciate any comments and suggestions from researchers using the macros.

## References


Brown, B. W., Jr. and M. Hollander, "Statistics a biomedical introduction", New York, John Wiley, 1977.

Cobb, G. and J. Gifford, Review of "Minitab handbook (Second edition)", Journal of the American Statistical Association 81 (1986), 1127-1129.

Gunter, B., "Macro for obtaining distinct values and replicate counts from a column of replicated data", Minitab Users' Group Newsletter 5 (1984), 1.

Miller, R. G., Jr., "Survival analysis", New York, John Wiley, 1981.

Ryan, B. F., B. L. Joiner and T. A. Ryan, Jr., "Minitab handbook (Second edition)", Boston, Duxbury Press, 1985.

## Appendix

We include here the listings of the two macros and a worked example from Miller (1981, pp. 49-50, pp. 203-205).

```
noecho
# macro km.geh for calculating the Kaplan-Meier ple and Gehan's test
# submitted by Andrew P. Soms, CMS, U of Wisconsin-Madison, 608-263-2681, and
# U of Wisconsin-Milwaukee, 414-229-4307, December 29, 1986
# input: data for sample 1 in c13, indicators (0 for censored
#        and 1 for completed) in c14, same for sample 2 in c16-c17
# scratch: c1-c12, c19, k1-k3
# output: c7-c10 K-M ple, simultaneous survival plots, c7(1)
#         Gehan's sum for sample 1, c8(1) standardized Gehan's
#         sum for sample 1, c9(1) 1-tailed P value, c10(1)
#         2-tailed P value, assuming normal approximation
# comments: All possible ties are allowed. The sorting algorithm is
#           an extension of the sorting algorithm by Bert Gunter,
#           Minitab Users' Group Newsletter, October 1984. The same
#           comments made there apply here - a program based on
#           column operations is substantially faster than one on
#           row operations.
#           Each group must contain at least 2 completed and
#           1 censored observation.
#           For further references see Statistics A Biomedical
#           Introduction, Brown and Hollander, or Survival
#           Analysis, Miller.
# program comments:
# 1/2 * smallest non-zero difference between ordered values added
# to censored observations to eliminate ties between completed
# and censored observations. This does not change K-M ple or
# Gehan's test. The ordered distinct values are put in c2,
# total replications in c3, indicator variables in c4, # in
# sample 1 in c5, # in sample 2 in c6.
let k1=count(c13)
set c15   # 1 for sample one
  k1(1)
  end
let k1=count(c16)
set c18   # 0 for sample two
  k1(0)
  end
copy c13-c15 c1-c3;
  use c14=0.
copy c13-c15 c4-c6;
  use c14=1.
copy c16-c18 c7-c9;
  use c17=0.
copy c16-c18 c10-c12;
  use c17=1.
stack c13 c16 c19
sort c19 c19
diff c19 c19
copy c19 c19;
  omit c19='*',0.
let k1=min(c19)/2   # 1/2 * smallest non-zero difference
let c1=c1+k1
let c7=c7+k1
stack (c1-c3) (c4-c6) (c7-c9) (c10-c12), (c1,c4,c5)
eras c2 c3 c6-c12
stack (c1,c4,c5) (c1,c4,c5), (c1,c4,c5)
copy c1 c2
sort c2,c4,c5 c2,c4,c5
stack (0,1,1) (c2,c4,c5) (c2,c4,c5)
diff c2 c2
```

```
copy c2 c2;
  omit c2='*'.
code (0)1 c2 c3
let c3=c3-c2
pars c3 c3
pars c4 c4
pars c5 c5
let k1=c3(count(c3))
let k2=c4(count(c4))
let k3=c5(count(c5))
let c3=sign(c2)*c3
stack c2 0 c2
let c4=sign(c2)*c4
let c5=sign(c2)*c5
copy c3 c3;
  omit c3=0.
copy c4 c4;
  omit c4=0.
copy c5 c5;
  omit c5=0.
stack 0 c3 k1 c3
stack c4 k2 c4
stack c5 k3 c5
diff c3 c3
copy c3 c3;
  omit c3='*'.
let c3=(c3+1)/2   # number of replications
diff c4 c4
copy c4 c4;
  omit c4='*'.
code (1:k2)1 c4 c4   # 1 if completed, 0 otherwise
diff c5 c5
copy c5 c5;
  omit c5='*'.
let c5=c5/2   # number in sample 1
let c6=c3-c5   # number in sample 2
copy c2 c2;
  omit c2=0.
pars c2 c2   # distinct values
# calculation of K-M ple for both samples
let k1=sum(c5)
copy c2,c4,c5 c7,c8,c9;
  omit c5=0.
stack 0 c9 c10
pars c10 c10
let k2=count(c10)
copy c10 c10;
  omit k2.
let c10=k1-c10   # risk set for sample 1
copy c7,c9,c10 c7,c9,c10;
  use c8=1.
let c8=(c10-c9)/c10
parp c8 c8   # K-M ple for sample 1
name c7 'com ob 1',c8 'K-M 1',c9 'reps'
print c7,c9,c8
let k1=sum(c6)
copy c2,c4,c6 c9,c10,c11;
  omit c6=0.
```

```
let k2=count(c12)
copy c12 c12;
  omit k2.
let c12=k1-c12   # risk set for sample 2
copy c9,c11,c12 c9,c11,c12;
  use c10=1.
let c10=(c12-c11)/c12
parp c10 c10   # K-M ple for sample 2
name c9 'com ob 2', c10 'K-M 2',c11 'reps'
print c9,c11,c10
echo
#
#
#
noecho
mplot c8 c7, c10 c9
echo
#
#
#
noecho
# calculation of Gehan sum of scores for sample 1, Mantel form
# of the variance and approximate P values
let c7=c3*c4
let k1=sum(c7)
pars c7 c9
stack 0 c7 c7
pars c7 c7
let k2=count(c7)
copy c7 c7;
  omit k2.
let c7=(c7-(k1-c7-c3))*c4
let c8=c3*(1-c4)
let k1=sum(c8)
pars c8 c8
let c8=(k1-c8)*c4
let c7=c7-c8   # Gehan's scores for completed observations, 0 otherwise
let c9=(1-c4)*c9   # Gehan's scores for censored observations, 0 otherwise
let c7=c7+c9   # Gehan's scores
let c8=c7**2
let k1=sum(c5)
let k2=sum(c6)
let k3=sqrt(sum(c3*c8)/((k1+k2)*(k1+k2-1))*k1*k2)   # Mantel variance
let k1=sum(c5*c7)
eras c7-c10
let c7(1)=k1
let k1=k1/k3
let c8(1)=k1
let k2=abso(k1)
cdf k1 k1;
  norm 0 1.
cdf k2 k2;
  norm 0 1.
let c9(1)=1-k1   # 1-tail P value that group 1 is better than 2
let c10(1)=2*(1-k2)   # 2-tail P value. This will be twice 1-tail P value
# only if Ge sum 1 ge 0, i.e., 1-tail P value le .5.
name c7 'Ge sum 1', c8 'st Ge s1', c9 '1-tail P', c10 '2-tail P'
print c7-c10
echo
end
```

```
noecho
# macro kmt.geh for calculating the Kaplan-Meier ple and Gehan's test
# submitted by Andrew P. Soms, CMS, U of Wisconsin-Madison, 608-263-2661, and
# U of Wisconsin-Milwaukee, 414-229-4307, December 29, 1986
# input: data for sample 1 in c13, indicators (0 for censored
#        and 1 for completed) in c14, same for sample 2 in c16-c17
# scratch: c1-c12, c19, k1-k4
# output: c7-c10 K-M ple, simultaneous survival plots, table
#         of ordered values and Gehan's scores, c7(1)
#         Gehan's sum for sample 1, c8(1) standardized Gehan's
#         sum for sample 1, c9(1) 1-tailed P value, c10(1)
#         2-tailed P value, assuming normal approximation
# comments: All possible ties are allowed. The sorting algorithm is
#           an extension of the sorting algorithm by Bert Gunter,
#           Minitab Users' Group Newsletter, October 1984. The same
#           comments made there apply here - a program based on
#           column operations is substantially faster than one on
#           row operations.
#           Each group must contain at least 2 completed and
#           1 censored observation.
#           For further references see Statistics A Biomedical
#           Introduction, Brown and Hollander, or Survival
#           Analysis, Miller.
# program comments:
# 1/2 * smallest non-zero difference between ordered values added
# to censored observations to eliminate ties between completed
# and censored observations. This does not change K-M ple or
# Gehan's test. The ordered distinct values are put in c2,
# total replications in c3, indicator variables in c4, # in
# sample 1 in c5, # in sample 2 in c6.
let k1=count(c13)
set c15   # 1 for sample one
  k1(1)
  end
let k1=count(c16)
set c18   # 0 for sample two
  k1(0)
  end
copy c13-c15 c1-c3;
  use c14=0.
copy c13-c15 c4-c6;
  use c14=1.
copy c16-c18 c7-c9;
  use c17=0.
copy c16-c18 c10-c12;
  use c17=1.
stack c13 c16 c19
sort c19 c19
diff c19 c19
copy c19 c19;
  omit c19='*',0.
let k1=min(c19)/2   # 1/2 * smallest non-zero difference
let k4=k1
let c1=c1+k1
let c7=c7+k1
stack (c1-c3) (c4-c6) (c7-c9) (c10-c12), (c1,c4,c5)
eras c2 c3 c6-c12
stack (c1,c4,c5) (c1,c4,c5), (c1,c4,c5)
```

```
stack [#,1,1] (c2,c4,c5) (c2,c4,c5)
diff c2 c2
copy c2 c2;
  omit c2='*'.
code (0)1 c2 c3
let c3=c3-c2
pars c3 c3
pars c4 c4
pars c5 c5
let k1=c3(count(c3))
let k2=c4(count(c4))
let k3=c5(count(c5))
let c3=sign(c2)*c3
stack c2 0 c2
let c4=sign(c2)*c4
let c5=sign(c2)*c5
copy c3 c3;
  omit c3=0.
copy c4 c4;
  omit c4=0.
copy c5 c5;
  omit c5=0.
stack 0 c3 k1 c3
stack c4 k2 c4
stack c5 k3 c5
diff c3 c3
copy c3 c3;
  omit c3='*'.
let c3=(c3+1)/2   # number of replications
diff c4 c4
copy c4 c4;
  omit c4='*'.
code (1:k2)1 c4 c4   # 1 if completed, 0 otherwise
diff c5 c5
copy c5 c5;
  omit c5='*'.
let c5=c5/2   # number in sample 1
let c6=c3-c5   # number in sample 2
copy c2 c2;
  omit c2=0.
pars c2 c2   # distinct values
# calculation of K-M ple for both samples
let k1=sum(c5)
copy c2,c4,c5 c7,c8,c9;
  omit c5=0.
stack 0 c9 c10
pars c10 c10
let k2=count(c10)
copy c10 c10;
  omit k2.
let c10=k1-c10   # risk set for sample 1
copy c7,c9,c10 c7,c9,c10;
  use c8=1.
let c8=(c10-c9)/c10
parp c8 c8   # K-M ple for sample 1
name c7 'com ob 1',c8 'K-M 1',c9 'reps'
print c7,c9,c8
let k1=sum(c6)
copy c2,c4,c6 c9,c10,c11;
  omit c6=0.
```

```
stack 0 c11 c12
pars c12 c12
let k2=count(c12)
copy c12 c12;
  omit k2.
let c12=k1-c12   # risk set for sample 2
copy c9,c11,c12 c9,c11,c12;
  use c10=1.
let c10=(c12-c11)/c12
parp c10 c10   # K-M plo for sample 2
name c9 'com ob 2', c10 'K-M 2',c11 'reps'
print c9,c11,c10
echo
#
#
#
noecho
mplot c8 c7, c10 c9
echo
#
#
#
noecho
# calculation of Gehan sum of scores for sample 1, Mantel form
# of the variance and approximate P values
let c7=c3*c4
let k1=sum(c7)
pars c7 c9
stack 0 c7 c7
pars c7 c7
let k2=count(c7)
copy c7 c7;
  omit k2.
let c7=(c7-(k1-c7-c3))*c4
let c8=c3*(1-c4)
let k1=sum(c8)
pars c8 c8
let c8=(k1-c8)*c4
let c7=c7-c8   # Gehan's scores for completed observations, 0 otherwise
let c9=(1-c4)*c9   # Gehan's scores for censored observations, 0 otherwise
let c7=c7+c9   # Gehan's scores
# print table. The next 6 command lines are the only way this
# program differs from the preceding one.
let c8=(c2-k4)*(1-c4)
let c9=c2*c4
let c2=c8+c9
name c5 'reps 1', c6 'reps 2'
name c2 'ob', c3 'to reps', c4 'indicat', c7 'G score'
print c2-c7
echo
#
#
#
noecho
let c8=c7**2
let k1=sum(c5)
let k2=sum(c6)
let k3=sqrt(sum(c3*c8)/((k1+k2)*(k1+k2-1))*k1*k2)   # Mantel variance
```

```
let c7(1)=k1
let k1=k1/k3
let c8(1)=k1
let k2=abso(k1)
cdf k1 k1;
  norm 0 1.
cdf k2 k2;
  norm 0 1.
let c9(1)=1-k1   # 1-tail P value that group 1 is better than 2
let c10(1)=2*(1-k2)   # 2-tail P value. This will be twice 1-tail P value
# only if Ge sum 1 ge 0, i.e., 1-tail P value le .5.
name c7 'Ge sum 1', c8 'st Ge s1', c9 '1-tail P', c10 '2-tail P'
print c7-c10
echo
end
```

```
MTB > # example from Miller, pp. 49-50 and 203-205
MTB > read 'rgm.1' c13 c14
     11 ROWS READ
 ROW    C13   C14

   1      9     1
   2     13     1
   3     13     0
   4     18     1


MTB > print c13-c14
 ROW    C13   C14

   1      9     1
   2     13     1
   3     13     0
   4     18     1
   5     23     1
   6     28     0
   7     31     1
   8     34     1
   9     45     0
  10     48     1
  11    161     0

MTB > read 'rgm.2' c16 c17
     12 ROWS READ
 ROW   C16   C17

   1     5     1
   2     5     1
   3     8     1
   4     8     1


MTB > print c16-c17
 ROW   C16   C17

   1     5     1
   2     5     1
   3     8     1
   4     8     1
   5    12     1
   6    16     0
   7    23     1
   8    27     1
   9    30     1
  10    33     1
  11    43     1
  12    45     1
```

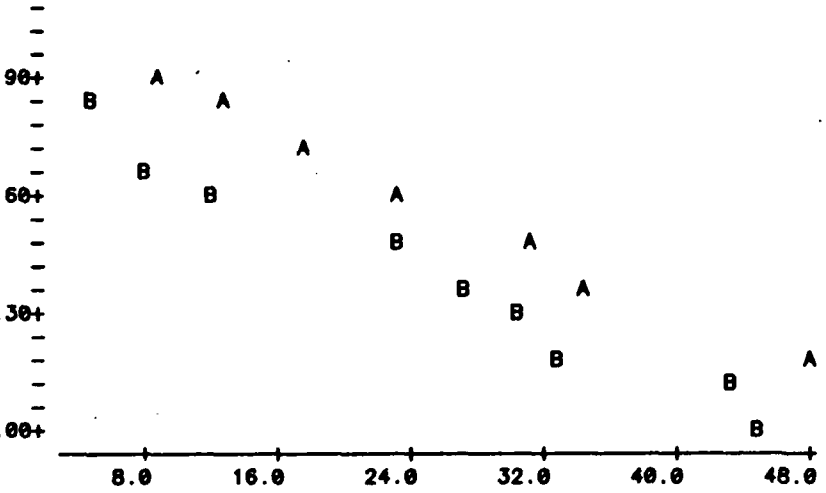
```
MTB > exec 'km.geh'
 ROW  com ob 1   reps      K-M 1

   1         9      1   0.909091
   2        13      1   0.818182
   3        18      1   0.715909
   4        23      1   0.613636
   5        31      1   0.490909
   6        34      1   0.368182
   7        48      1   0.184091

 ROW  com ob 2   reps      K-M 2

   1         5      2   0.833333
   2         8      2   0.666667
   3        12      1   0.583333
   4        23      1   0.486111
   5        27      1   0.388889
   6        30      1   0.291667
   7        33      1   0.194444
   8        43      1   0.097222
   9        45      1   0.000000

MTB > #
MTB > #
MTB > #
```




```
MTB > #
MTB > #
MTB > #
```

```
 ROW  Ge sum 1  st Ge s1    1-tail P    2-tail P

   1        50   1.65567   0.0488948   0.0977895

MTB > end
MTB > outfile
```

```
MTB > # example from Miller, pp. 49-50 and 203-205
MTB > read 'rgm.1' c13 c14
     11 ROWS READ
 ROW    C13   C14

   1      9     1
   2     13     1
   3     13     0
   4     18     1


MTB > print c13-c14
 ROW    C13   C14

   1      9     1
   2     13     1
   3     13     0
   4     18     1
   5     23     1
   6     28     0
   7     31     1
   8     34     1
   9     45     0
  10     48     1
  11    161     0

MTB > read 'rgm.2' c16-c17
     12 ROWS READ
 ROW   C16   C17

   1      5     1
   2      5     1
   3      8     1
   4      8     1


MTB > print c16-c17
 ROW   C16   C17

   1      5     1
   2      5     1
   3      8     1
   4      8     1
   5     12     1
   6     16     0
   7     23     1
   8     27     1
   9     30     1
  10     33     1
  11     43     1
  12     45     1
```

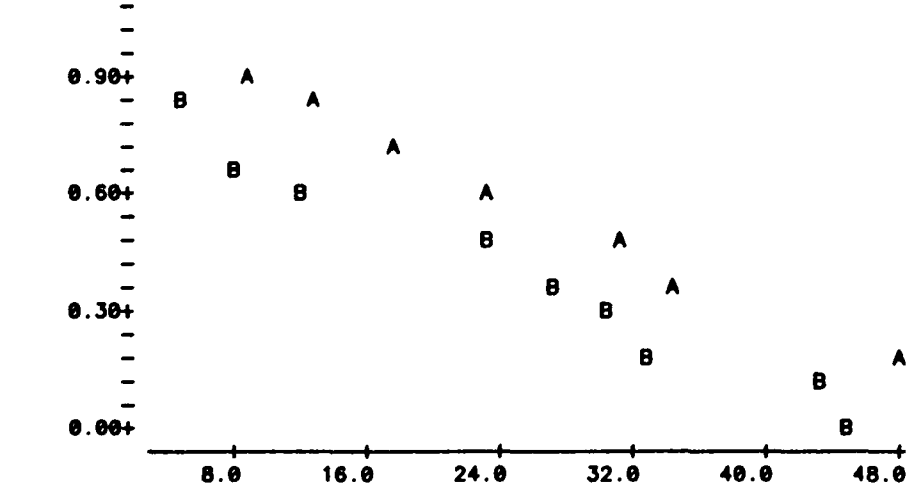
```
MTB > exec 'kmt.geh'
 ROW  com ob 1   reps      K-M 1

   1         9      1   0.909091
   2        13      1   0.818182
   3        18      1   0.715909
   4        23      1   0.613636
   5        31      1   0.490909
   6        34      1   0.368182
   7        48      1   0.184091

 ROW  com ob 2   reps      K-M 2

   1         5      2   0.833333
   2         8      2   0.666667
   3        12      1   0.583333
   4        23      1   0.486111
   5        27      1   0.388889
   6        30      1   0.291667
   7        33      1   0.194444
   8        43      1   0.097222
   9        45      1   0.000000

MTB > #
MTB > #
MTB > #
```




```
MTB > #
MTB > #
MTB > #
```

| ROW | ob | to reps | Indicat | reps 1 | reps 2 | G score |
|---|---|---|---|---|---|---|
| 1 | 5 | 2 | 1 | 0 | 2 | -21 |
| 2 | 8 | 2 | 1 | 0 | 2 | -17 |
| 3 | 9 | 1 | 1 | 1 | 0 | -14 |
| 4 | 12 | 1 | 1 | 0 | 1 | -12 |
| 5 | 13 | 1 | 1 | 1 | 0 | -10 |
| 6 | 13 | 1 | 0 | 1 | 0 | 7 |
| 7 | 16 | 1 | 0 | 0 | 1 | 7 |
| 8 | 18 | 1 | 1 | 1 | 0 | -6 |
| 9 | 23 | 2 | 1 | 1 | 1 | -3 |
| 10 | 27 | 1 | 1 | 0 | 1 | 0 |
| 11 | 28 | 1 | 0 | 1 | 0 | 11 |
| 12 | 30 | 1 | 1 | 0 | 1 | 3 |
| 13 | 31 | 1 | 1 | 1 | 0 | 5 |
| 14 | 33 | 1 | 1 | 0 | 1 | 7 |
| 15 | 34 | 1 | 1 | 1 | 0 | 9 |
| 16 | 43 | 1 | 1 | 0 | 1 | 11 |
| 17 | 45 | 1 | 1 | 0 | 1 | 13 |
| 18 | 45 | 1 | 0 | 1 | 0 | 17 |
| 19 | 48 | 1 | 1 | 1 | 0 | 16 |
| 20 | 161 | 1 | 0 | 1 | 0 | 18 |

```
MTB > #
MTB > #
MTB > #
```

| ROW | Ge sum 1 | st Ge s1 | 1-tail P | 2-tail P |
|---|---|---|---|---|
| 1 | 50 | 1.65567 | 0.0488948 | 0.0977895 |

```
MTB > end
MTB > outfile
```

END
DATED
FILM
8-88
DTIC