Introduction

The Jonckheere-Terpstra test (Terpstra, 1952 and Jonckheere, 1954) is a nonparametric method for evaluating \( k > 2 \) comparison (e.g. treatment) groups under the null hypothesis of equal outcomes among the groups against the one-sided alternative that outcomes follow a definite ordering, e.g. \( T_1 \leq T_2 \leq \ldots \leq T_k \), with at least one strict inequality. This method may therefore be regarded as an analysis of trend or dose response. If the treatments are \textit{a priori} appropriately ordered and the outcomes are assumed to follow this ordering, then this test is more powerful than the nonparametric Kruskal-Wallis test or a standard one-way Analysis of Variance, for which the alternative hypothesis does not take treatment ordering into account.

The original test did not allow for tied outcomes. However, later work has made analysis in the presence of ties possible. This paper reviews the Jonckheere-Terpstra (henceforth J-T) test and presents a general SAS Macro for analyzing \( n_i \) outcomes (\( n_1, \ldots, n_k \) not necessarily equal, with possible ties) in \( k > 2 \) treatment groups.

The test

The J-T test is simple in concept but can be difficult to use without a computer program. For purposes of this presentation, it is assumed that treatments increase in a left-to-right direction. The data within each treatment group are sorted in an ascending order. Within the \( k(k-1)/2 \) pairs \( i, j \) of treatment groups individual outcomes are compared. Pairs of outcomes where \( x_{i\alpha} < x_{j\beta} \) are scored as 1, and pairs where \( x_{i\alpha} > x_{j\beta} \) are scored as 0. Pairs where \( x_{i\alpha} = x_{j\beta} \) are scored as 1/2. Scores within each treatment pair are summed to form Mann-Whitney statistics \( U_{ij} \). The \( U_{ij} \)'s are then summed to form statistic J. The standardized J-T test statistic is then \( Z = (J - E_0(J))/\sqrt{VAR_0(J)} \), where \( E_0(J) \) and \( VAR_0(J) \) are the expected value and variance of J, respectively, under the null hypothesis:

\[ Z = \frac{J - E_0(J)}{\sqrt{VAR_0(J)}} \]
\[ E_o(J) = \frac{(N^2 - \Sigma n_i^2)}{4}, \]
\[ VAR_o(J) = \left[ \frac{\left( N(N-1)(2N+5) - \Sigma n_i(n_i-1)(2n_i+5) - \Sigma d_j(d_j-1)(2d_j+5) \right)}{72} \right] \]
\[ + \frac{\left[ \Sigma n_i(n_i-1)(n_i-2) \right] \left[ \Sigma d_j(d_j-1)(d_j-2) \right]}{(36N(N-1)(N-2))} \]
\[ + \frac{\left[ \Sigma n_i(n_i-1) \right] \left[ \Sigma d_j(d_j-1) \right]}{(8N(N-1))}. \]

Here, \( N \) is the total number of outcomes, \( n_i \) is the number of outcomes in treatment group \( i \), and \( d_j \) is the number of occurrences of outcome \( j \).

The program

The program to calculate the standardized J-T statistic makes extensive use of BASE SAS procedures (namely PROC MEANS and PROC FREQ) and MACRO processing. First, the number of treatment groups \( k \) is determined. Then the number of outcomes \( n_i \) within each treatment is found and used in calculating \( E_o(J) \) and \( VAR_o(J) \). The frequency of each outcome \( d_j \) across all treatment groups is determined and used in calculating \( VAR_o(J) \). \( J \) itself is calculated by setting up two loops called OUTER and INNER. OUTER goes from treatment group 1 through \( k-1 \), while INNER goes from treatment group 2 through \( k \). Within each index of OUTER pairwise outcome comparisons are made within each index of INNER and Mann-Whitney statistics \( U_{ij} \) are found and added to \( J \). Finally the standardized J-T statistic \( Z \) is obtained and the \( p \)-value calculated.

The program is written assuming that treatment outcomes follow an ascending ordering. If the reverse is true, i.e. \( T_1 \geq T_2 \geq \ldots \geq T_k \) then either the ordering of the treatments is reversed to \( (k+1) - i \), \( i=1, \ldots, k \), or an adjustment can be made to the data as shown in example 2.

The program code can be found in Appendix 1.
Example 1

The following is taken from Jonckheere (1954). An experiment yields four outcomes in each of four independent groups. A priori, the groups were assumed to be in ascending order. The null hypothesis is that the four samples come from the same population against the alternative that outcomes in the four groups are in order of increasing value. The data are:

<table>
<thead>
<tr>
<th></th>
<th>I (1)</th>
<th>II (2)</th>
<th>III (3)</th>
<th>IV (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>19</td>
<td>21</td>
<td>40</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>61</td>
<td>99</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>80</td>
<td>100</td>
<td>151</td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>129</td>
<td>149</td>
<td>160</td>
</tr>
</tbody>
</table>

Loop OUTER begins at treatment 1 and ends at 3 while loop INNER begins at treatment 2 and ends at 4. \( U_{12} = 11 \) (19<21, 20<21, 19<61, 20<61, 60<61, 19<80, 20<80, 60<80, 19<129, 20<129, and 60<129). Similarly, \( U_{13} = 12 \), \( U_{14} = 13 \), \( U_{23} = 11 \), \( U_{24} = 12 \), and \( U_{34} = 12 \). \( J = \Sigma U_{ij} = 11+12+13+11+12+12+12 = 71 \), \( E_0(J) = 48 \), \( \text{VAR}_0(J) = 10.708 \), and \( Z = (J-E_0(J)) / \text{SQRT(VAR}_0(J)) = 2.148 \), resulting in \( p = 0.016 \).

Note: in this example there were no ties and \( n_1 = n_2 = n_3 = n_4 \).

Example 2

Three treatment groups, 0 dose active (placebo), low dose active and high dose active have been administered to patients over a period of time to assess change in a clinical endpoint. In this study negative changes, or decreases from baseline, are considered as favorable outcomes. The data follow:
We now have a problem with unequal ns and ties among the outcomes. Also to get the data into the appropriate order so that $T_1 \leq T_2 \leq T_3$ each outcome was multiplied by -1. The program was run yielding the following results: $J=482.5$, $E_0(J)=266$, $VAR_0(J)=40.202$, $Z=5.385$, with $p<.001$. The inference is a higher dose results in a greater decrease from baseline.

Note that if the data were analyzed without premultiplying each outcome by -1, $E_0(J)$, $VAR_0(J)$ and $p$ would remain unchanged. However, $J=49.5$ and $Z=-5.385$, and the alternative hypothesis would need to be revised to adjust for direction of treatment-outcomes.

Conclusions

The Jonckheere-Terpstra test is a powerful method for testing ordered alternatives to the null hypothesis of equal treatment outcomes. Unfortunately, while the test is fairly simple in concept, it can quickly become difficult to apply without computer programming. The macro program presented here is a such a tool. It is generalized to allow unequal sample sizes and tied data for any $k>2$ comparison groups. It is easy to use and with a simple up-front change can be applied to alternatives going in either direction.
References


*BASE SAS is a registered trademark of SAS Institute Inc., Cary, NC.
Example 1
Jonckheere-Terpstra Test for Ordered Alternatives
Data Source: Jonckheere A.R., A Distribution-Free k-Sample Test Against Ordered Alternatives, Biometrica, v. 41, 1954, pp.133-143.

<table>
<thead>
<tr>
<th>OBS</th>
<th>J</th>
<th>DUMMY</th>
<th>U12</th>
<th>U13</th>
<th>U14</th>
<th>U23</th>
<th>U24</th>
<th>U34</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>7</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>59</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>59</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>59</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>59</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>71</td>
<td>1</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBS</th>
<th>J</th>
<th>EXPECTED</th>
<th>SND</th>
<th>Z</th>
<th>P.VAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71</td>
<td>48</td>
<td>10.7083</td>
<td>2.148</td>
<td>0.016</td>
</tr>
</tbody>
</table>
### Example 2

<table>
<thead>
<tr>
<th>OBS</th>
<th>J</th>
<th>OBS</th>
<th>U12</th>
<th>U13</th>
<th>U23</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>1</td>
<td>3.0</td>
<td>11.5</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>1</td>
<td>9.5</td>
<td>24.5</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>0.0</td>
<td>1</td>
<td>19.0</td>
<td>38.5</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
<td>1</td>
<td>30.5</td>
<td>52.5</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>0.0</td>
<td>1</td>
<td>44.5</td>
<td>66.5</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>0.0</td>
<td>1</td>
<td>58.5</td>
<td>80.5</td>
<td>0.0</td>
</tr>
<tr>
<td>7</td>
<td>0.0</td>
<td>1</td>
<td>72.5</td>
<td>94.5</td>
<td>0.0</td>
</tr>
<tr>
<td>8</td>
<td>0.0</td>
<td>1</td>
<td>86.5</td>
<td>108.5</td>
<td>0.0</td>
</tr>
<tr>
<td>9</td>
<td>0.0</td>
<td>1</td>
<td>100.5</td>
<td>122.5</td>
<td>0.0</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>1</td>
<td>114.5</td>
<td>136.5</td>
<td>0.0</td>
</tr>
<tr>
<td>11</td>
<td>0.0</td>
<td>1</td>
<td>128.5</td>
<td>150.5</td>
<td>0.0</td>
</tr>
<tr>
<td>12</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>0.0</td>
</tr>
<tr>
<td>13</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>7.5</td>
</tr>
<tr>
<td>14</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>12.0</td>
</tr>
<tr>
<td>15</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>20.5</td>
</tr>
<tr>
<td>16</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>40.5</td>
</tr>
<tr>
<td>17</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>53.0</td>
</tr>
<tr>
<td>18</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>62.5</td>
</tr>
<tr>
<td>19</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>78.5</td>
</tr>
<tr>
<td>20</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>92.0</td>
</tr>
<tr>
<td>21</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>105.5</td>
</tr>
<tr>
<td>22</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>119.5</td>
</tr>
<tr>
<td>23</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>133.5</td>
</tr>
<tr>
<td>24</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>147.5</td>
</tr>
<tr>
<td>25</td>
<td>107.0</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>161.5</td>
</tr>
<tr>
<td>26</td>
<td>402.5</td>
<td>1</td>
<td>142.5</td>
<td>164.5</td>
<td>175.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OBS</th>
<th>J</th>
<th>EXPECTED</th>
<th>STDW</th>
<th>I</th>
<th>P_VAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>402.5</td>
<td>286</td>
<td>40.2020</td>
<td>5.385</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Appendix 1

data enter;

********** Bring in your data set here:
run;
proc sort data=enter out=enter;
by trt;
run;

data ntrts;
set enter;
by trt;
if first.trt;
run;
proc means n noprint data=ntrts;
  id dummy;
  var trt;
  output out=ntrts n=ntrt; * Number of unique treatments;
run;

data ntrts;
set ntrts;
call symput('ntrt', ntrt); * Used to define array size;
run;
%macro jt_test (start, stop);
proc means data=enter n noprint;
  id dummy;
  by trt;
  var resp;
  output out=trtcnts n=nObs; * Number of obs per treatment;
run;
data trtcnts;
set trtcnts;
  by trt;
  nprodl=nObs*(nObs-1)*((2*nObs)+5) ; * N(N-1)(2N+5);
  nprod2=nObs*(nObs-1)*(nObs-2) ; * N(N-1)(N-2);
  nprod3=nObs*(nObs-1) ; * N(N-1);
  nobssq=nObs*nObs ; * N**2;
run;
* Dataset TRTSUMS calculated below contains sums of obs by treatments, as well as sums over all treatments of the products used to derive the variance of the Jonkheere-Terpstra test:
proc means data=trtcnts sum noprint;
  id dummy;
  var nObs nprodl nprod2 nprod3 nobssq;
  output out=trtsums sum=sumobs sumprodl sumprod2 sumprod3 sumobssq;
run;
data trtsums;
set trtsums;
* The following products using overall N across all treatments are used in calculating variance of Jonckheere-Terpstra test:
  sumobs1=sumobs*(sumobs-1)*((2*sumobs)+5) ; * N(N-1)(2N+5);
  sumobs2=sumobs*(sumobs-1)*(sumobs-2) ; * N(N-1)(N-2);
  sumobs3=sumobs*(sumobs-1) ; * N(N-1);
  expected=((sumobs*sumobs)-sumobssq)/4; * Expected value of the sum of Mann-Whitney U statistics;
run;
**** Next find the frequency of each outcome regardless of treatment:
proc freq data=enter;
  tables resp / out=datafreq noprint;
run;

42
data datafreq(drop=resp percent);
  set datafreq;
  cntprod1=count*(count-l)*(2*count+5); * di(di-l)(2di+5);
  cntprod2=count*(count-l)*(count-2); * di(di-l)(di+2);
  cntprod3=count*(count-l); * di(di-l);
  dummy=l;
run;

proc means data=datafreq sum noprint;
  id dummy;
  var cntprod1 cntprod2 cntprod3;
  output out=cntsums sum=cntsum1 cntsum2 cntsum3; * These are the sums of products of frequencies of unique data values.;
run;

data products;
  merge trtcnts trtsums cntsums;
  by dummy;
  if first.dummy;
  * Next calculate the Standard Deviation used to find p-value.
  It is corrected for ties.;
  stdw=sqrt((sumObs1-sumprod1-cntsum1)/72) +
  ((sumprod2*cntsum2)/(36*sumobs2)) +
  ((sumprod3*cntsum3)/(8*sumobs3));
run;

%do s = &start %to &stop;
  data trt&s;
    set enter;
    if trt = &s;
      if resp=. then delete;
      rename resp resp&s;
    keep trt resp dummy;
run;
%end;

%let outend=%eval (&stop-1); * Outer loop goes from 1 to k-1;

data MannWhit;
  retain j 0;
  keep dummy j;
  %do outer=#start %to #outend; * Focus on each obs in OUTER loop;
    do number=1 to tot&outer;
      set trt&outer point=number nObs=tot&outer;
      %let instart=%eval (#outer+1); * Inner loop goes from outer loop lower limit+1 to k;
    %do inner=#instart %to #stop; * Focus on each obs in INNER loop;
      keep u&outer&inner;
      do num=1 to tot&inner;
        set trt&inner point=numb nObs=tot&inner;
        retain u&outer&inner 0;
        if (resp&outer > resp&inner) then
          u&outer&inner=u&outer&inner + 0; * If n1>n2, score=0;
        else if resp&outer=resp&inner = 0 then
          u&outer&inner=u&outer&inner + .5; * Tied observations;
        else if (resp&outer < resp&inner) then
          u&outer&inner=u&outer&inner + 1; * If n1<n2, score=1;
        **** J is calculated below as a running total as loops increment;
        if number=tot&outer & numb=tot&inner then j=j+u&outer&inner;
      end;
    end;
  end;
%end;
stop;
run;
proc print data=mannwhit;
run;

data mannwhit(drop=dummy);
merge mannwhit products;
   by dummy;
      if last.dummy;
         Z=(j-expected)/stdw;
         p_val=1-probnorm(abs(Z));
run;

proc print data=mannwhit;
   var j expected stdw Z p_val;
run;

title; run;

%mend j_t_test;

%j_t_test(1, &ntrt)