Jonckheere-Terpstra test

Tao Xu

STAT 6500
Jonckheere-Terpstra test
(Jonckheere trend test)

• Nonparametric test

• Similar to the Kruskal-Wallis test

• More statistical power in ordered alternative
Hypothesis

- $H_0: \theta_1 = \theta_2 = \theta_3 = \ldots = \theta_k$

- $H_A: \theta_1 \leq \theta_2 \leq \theta_3 \leq \ldots \leq \theta_k$ (with at least one strict inequality)

- $k > 2$

- $\Theta_i$ is the population median for $i$th population
\[ Z = \frac{U - E(U)}{\sqrt{\text{Var}(U)}} \]

For large \( N \) and the individual \( n_i \) not too small, the distribution of test statistic \( Z \) is approximately standard normal.

Where

\( U_{xy} \) is the number of observations in group \( y \) that are greater than each observation in group \( x \).

\[ E(U) = \frac{N^2 - \sum i n_i^2}{4} \]

\( N \): total sample size

\( n_i \): sample number in each group

\[ \text{Var}(U) = \frac{N^2(2N+3) - \sum i [n_i^2(2n_i+3)]}{72} \]
Examples

• Hinkley (1989) gives braking distances taken by motorists to stop when travelling at various speeds. A subset of his data is shown as follow.

<table>
<thead>
<tr>
<th>Speed (mph)</th>
<th>Braking distances (feet)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>48</td>
</tr>
<tr>
<td>25</td>
<td>33  59  48  56</td>
</tr>
<tr>
<td>30</td>
<td>60  101  67</td>
</tr>
<tr>
<td>35</td>
<td>85  107</td>
</tr>
</tbody>
</table>
Hypothesis

• $H_0$: braking distance is independent of initial speed.
• $H_A$: braking distance increases when initial speed increases.

• Isn’t that obvious?
Test statistic calculation

- $U_{xy} = U_{12} + U_{13} + U_{14} + U_{23} + U_{24} + U_{34} = 2.5 + 3 + 2 + 12 + 8 + 5 = 32.5$
- $E(U) = 17.5$
- $Var(U) = 27.9$
- $Z = 2.84$
- One-tail $p = 0.0022$
**SAS code**

```sas
data breaking;
input  speed  distance @@;
cards;
20  48
25  33
25  59
25  48
25  56
30  60
30 101
30  67
30  85
35  85
35 107
;
proc npar1way wilcoxon data=breaking;
class  speed;
var  distance;
title1 'Kruskal-Wallis Test';
run;
proc freq;
table speed*distance / jt nopercent nocol norow;
run;
```
Outputs

Not significant result in Kruskal-Wallis Test. However, significant result in Jonckheere-Terpstra Test.

Kruskal-Wallis Test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>7.2599</td>
</tr>
<tr>
<td>DF</td>
<td>3</td>
</tr>
<tr>
<td>Pr &gt; Chi-Square</td>
<td>0.0641</td>
</tr>
</tbody>
</table>

Jonckheere-Terpstra Test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>32.5000</td>
</tr>
<tr>
<td>Z</td>
<td>2.8489</td>
</tr>
<tr>
<td>One-sided Pr &gt; Z</td>
<td>0.0022</td>
</tr>
<tr>
<td>Two-sided Pr &gt;</td>
<td>Z</td>
</tr>
</tbody>
</table>
> install.packages("clinfun")
> library(clinfun)
> speed20<-c(48)
> speed25<-c(33,59,48,56)
> speed30<-c(60,101,67)
> speed35<-c(85,107)
> speed<-c(speed20,speed25,speed30,speed35)
> distance<-c(rep(1,1),rep(2,4),rep(3,3),rep(4,2))
> result<-kruskal.test(speed,distance)
> result

> pieces<-list(speed20, speed25, speed30, speed35)
> n<-c(1,4,3,2)
> grp<-as.ordered(factor(rep(1:length(n),n)))
> jonckheere.test(unlist(pieces),grp,alternative="increasing")
Kruskal-Wallis rank sum test
data:  speed and distance
Kruskal-Wallis chi-squared = 7.2599, df = 3, p-value = 0.06406

Jonckheere-Terpstra test
data:
JT = 32.5, p-value = 0.002263
alternative hypothesis: increasing
Conclusion

• There is clear evidence that increasing speed increases braking distance.
• When the speed and braking distance are relevant, the Jonckheere-Terpstra test is generally more powerful than Kruskal-Wallis test (higher probability that the test will reject the $H_0$ when the $H_A$ is true).
THANK YOU