Test 50  The median test of two populations

Object
To test if two random samples could have come from two populations with the same frequency distribution.

Limitations
The two samples are assumed to be reasonably large.

Method
The median of the combined sample of \( n_1 + n_2 \) elements is found. Then, for each series in turn, the number of elements above and below this median can be found and entered in a \( 2 \times 2 \) table of the form:

<table>
<thead>
<tr>
<th></th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left of median</td>
<td>( a )</td>
<td>( b )</td>
<td>( a + b )</td>
</tr>
<tr>
<td>Right of median</td>
<td>( c )</td>
<td>( d )</td>
<td>( c + d )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_1 = a + c )</td>
<td>( n_2 = b + d )</td>
<td>( N = n_1 + n_2 )</td>
</tr>
</tbody>
</table>

The test statistic is

\[
\chi^2 = \frac{|ad - bc| - \frac{1}{2}N^2}{(a+b)(a+c)(b+d)(c+d)}
\]

If this value exceeds the critical value obtained from \( \chi^2 \) tables with one degree of freedom, the null hypothesis of the same frequency distribution is rejected.

Example
A housing officer has data relating to residents’ assessment of their housing conditions in a small, isolated estate. Half of the houses in the estate are maintained by one maintenance company and the other half by another company. Do the repair regimes of the two companies produce similar results from the residents? Samples of 15 residents are taken from each half. The calculated chi-squared value is 0.53, which is less than the tabulated value of 3.84. The housing officer does not reject the null hypothesis and concludes that the two maintenance companies produce similar results from their repair regimes.

Numerical calculation
\( a = 9, b = 6, c = 6, d = 9 \)
\( a + b = a + c = b + d = c + d = 15 \)
\( n_1 = 15, n_2 = 15, N = 30 \)
\( \chi^2 = \frac{|9^2 - 6^2| - \frac{1}{2} \times 30}{15 \times 15 \times 15 \times 15} = \frac{8}{15} = 0.53 \)
\( \chi^2_{1,0.05} = 3.84 \) [Table 5]
Do not reject the null hypothesis.
Test 51  The median test of $K$ populations

**Object**

To test if $K$ random samples could have come from $K$ populations with the same frequency distribution.

**Limitations**

The $K$ samples are assumed to be reasonably large – say, greater than 5.

**Method**

The $K$ samples are first amalgamated and treated as a single grand sample, of which the median is found. Then, for each of the $K$ samples, the number of elements above and below this median can be found. These can be arranged in the form of a $2 \times K$ table and then a $\chi^2$-test can be carried out.

<table>
<thead>
<tr>
<th>Sample</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>$j$</th>
<th>...</th>
<th>$K$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above median</td>
<td>$a_{11}$</td>
<td>$a_{12}$</td>
<td>...</td>
<td>$a_{1j}$</td>
<td>...</td>
<td>$a_{1K}$</td>
<td>A</td>
</tr>
<tr>
<td>Below median</td>
<td>$a_{21}$</td>
<td>$a_{22}$</td>
<td>...</td>
<td>$a_{2j}$</td>
<td>...</td>
<td>$a_{2K}$</td>
<td>B</td>
</tr>
<tr>
<td>Total</td>
<td>$a_1$</td>
<td>$a_2$</td>
<td>...</td>
<td>$a_j$</td>
<td>...</td>
<td>$a_K$</td>
<td>$N$</td>
</tr>
</tbody>
</table>

In this table, $a_{1j}$ represents the number of elements above the median and $a_{2j}$ the number of elements below the median in the $j$th sample ($j = 1, 2, \ldots, K$). Expected frequencies are calculated from

$$e_{1j} = \frac{A a_j}{N} \quad \text{and} \quad e_{2j} = \frac{B a_j}{N}.$$  

The test statistic is

$$\chi^2 = \sum_{j=1}^{K} \frac{(a_{1j} - e_{1j})^2}{e_{1j}} + \sum_{j=1}^{K} \frac{(a_{2j} - e_{2j})^2}{e_{2j}}.$$  

This is compared with a critical value from Table 5 with $K - 1$ degrees of freedom. The null hypothesis that the $K$ populations have the same frequency distribution is rejected if $\chi^2$ exceeds the critical value.

**Example**

The housing officer in test 50 has a larger estate which is maintained by five maintenance companies. He has sampled the residents receiving maintenance from each company in proportion to the number of houses each company maintains. The officer now produces a chi-squared value of 0.2041. Do the five maintenance companies differ in their effect on resident’s assessment? The tabulated chi-squared value is 9.49, so the officer concludes that the standards of maintenance are the same.
**Numerical calculation**

<table>
<thead>
<tr>
<th>Sample</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above median</td>
<td>20</td>
<td>30</td>
<td>25</td>
<td>40</td>
<td>30</td>
<td>145</td>
</tr>
<tr>
<td>Below median</td>
<td>25</td>
<td>35</td>
<td>30</td>
<td>45</td>
<td>32</td>
<td>167</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>65</td>
<td>55</td>
<td>85</td>
<td>62</td>
<td>312</td>
</tr>
</tbody>
</table>

\[ e_{11} = \frac{145 \times 45}{312} = 20.91 \quad e_{21} = \frac{167 \times 45}{312} = 24.08 \]

\[ e_{12} = 30.21 \quad e_{22} = 34.79 \]
\[ e_{13} = 25.56 \quad e_{23} = 29.44 \]
\[ e_{14} = 39.50 \quad e_{24} = 45.50 \]
\[ e_{15} = 28.81 \quad e_{25} = 33.19 \]

\[ \chi^2 = 0.0396 + 0.0015 + 0.0123 + 0.0063 + 0.0491 + 0.0351 + 0.0013 + 0.0107 + 0.0055 + 0.0427 \]
\[ \quad = 0.2041 \]

\[ \chi^2_{4;0.05} = 9.49 \text{ [Table 5].} \]

Hence do not reject the null hypothesis.