The Median Test – Analysis of k-Between-Group Data with a Quantitative Response Variable

Application: To compare the medians of a quantitative variable obtained from 2 or more groups. Thus, it is applied in the same data situation as an ANOVA for independent samples, except that it is used when the data are either importantly non-normally distributed, the measurement scale of the dependent variable is ordinal (not interval or ratio), or from a too-small sample. It is important to remember the null hypothesis for this test, and to differentiate it from the nulls for the t-test, Mann-Whitney U, and Kruskal-Wallis test.

H0: The populations represented by the k conditions (groups, samples) have the same median on the quantitative response variable.

To reject H0: is to say that the populations have different medians.

The data: This analysis involves the grouping variable **chain** (1 = chain store, 2 = privately owned store, 3 = coop owned store) and the response variable **fishnum** (number of fish on display). Below are the scores for the 12 stores (chain, fishnum).

        3,32   3,41   3,31   3,38   1,21   1,13   2,17   1,22   2,24   1,11   2,17   1,20

Research Hypothesis: The data for the analysis come from the Pet Shop database. The researcher hypothesized that Coop stores would have the most fish on display, chain stores would display the least, and private pet stores would display an intermediate amount.

H0: for this analysis : The three different types of pet shops have the same median number of fish displayed.

Step 1 Rearrange the data from lowest to highest score while keeping track of group membership.

Group ➜	1	1	2	2	1	1	1	2	3	3	3	3
Fishnum ➜	11	13	17	17	20	21	22	24	31	32	38	41

Step 2 Compute the median score for the total sample (ignoring group membership).

   - If there is an odd number of data values, the median is the middle-most value.

   - If there is an even number of data values, the median is the average of the middle-most two.

    This sample has 12 cases, so the median is the average of the values of the
    middle **two** scores.   ➜ (21 + 22) / 2 = 21.5

Step 3 Assemble the data for analysis. Organize the data into a contingency table so that each score is assigned to a cell defined by the group to which it belongs and whether that score exceeds the median of the combined sample or not (is equal to or below the combined median). The number of stores that fall into each cell is called the obtained frequency (**of**) for that cell.

	Group Membership		
Score	Chain Store	Private Store	Coop Store
	of	of	of
Above Median	1	1	4
Below Median	4	2	0

Step 4 Compute the row totals (sum across the values on the same row), column totals (sum across the values on the same column) and overall totals of the observed frequencies.

```
                                    Group Membership

       Score        Chain Store      Private Store      Coop Store        Row Totals

                 ┌───────────────┬───────────────────┬──────────────┐
Above Median     │       1       │         1         │       4      │          6
                 ├───────────────┼───────────────────┼──────────────┤
Below Median     │       4       │         2         │       0      │          6
                 └───────────────┴───────────────────┴──────────────┘        Overall Total
       Column Totals    5                  3                  4                    12
```

Step 5 Compute the expected frequency (**ef**) of each cell. This expected frequency is computed as the product of the row total and the column total for that cell, divided by the overall total. For example, the upper left-hand cell (the num ber of chain stores that displayed more than the median number of fish) has an expected frequency of:

$$ef \quad = \quad \frac{\text{row total} * \text{column total}}{\text{overall total}} \quad = \quad \frac{6 * 5}{12} \quad = \quad 2.5$$

```
                                Group Membership

Score              Chain Store       Private Store       Coop Store

                    of    ef           of    ef           of    ef
                 ┌───────────────┬───────────────────┬──────────────┐
Above Median     │    1    2.5   │     1    1.5      │   4    2.0    │
                 ├───────────────┼───────────────────┼──────────────┤
Below Median     │    4    2.5   │     2    1.5      │   0    2.0    │
                 └───────────────┴───────────────────┴──────────────┘
```

Step 6 Compute Chi-Square (apply the following formula to each cell, and sum the results).

$$X^2 \quad = \quad \Sigma \frac{(of-ef)^2}{ef} \quad = \quad \frac{(1-2.5)^2}{2.5} + \frac{(1-1.5)^2}{1.5} + \frac{(4-2)^2}{2.0} + \frac{(4-2.5)^2}{2.5} + \frac{(2-1.5)^2}{1.5} + \frac{(0-2)^2}{2.0}$$

$$= .9 \; + \; .17 \; + \; 2 \; + \; .9 \; + \; .17 \; + \; 2 \; = \; 6.14$$

Step 7 Compute the degrees of freedom (df)

$$df \; = \; (\text{number of columns } - 1) * (\text{number of rows } - 1) \; = \; (3-1) * (2-1) \; = \; 2 * 1 \; = \; 2$$

Step 8 Use the table of X^2 values to determine the critical X^2 value for df = 2 and p= .05

$$X^2(df=2, \; p=.05) \; = \; 5.991$$

Step 9 Compare the obtained X² and critical X², and determine whether or not there is a statistically significant difference among the group medians

    -- if the obtained X² is less than the critical X², then retain the null hypothesis -- conclude that there is no difference between the medians of the populations represented by the samples

    -- if the obtained X² is greater than the critical X², then reject the null hypothesis -- conclude that there is a difference among the medians of the populations represented in the samples.

       For the example data, we would decide to reject the null hypothesis, because the obtained X² value of 6.14 is larger than the critical Chi-square value of 5.991.

Step 10 IF you reject the null hypothesis, determine whether the pattern of the medians completely supports, partially supports, or does not support the research hypothesis.

    -- IF you reject the null hypothesis, AND if the pattern the medians agrees exactly with the research hypothesis, then the research hypothesis is completely supported.

    -- IF you reject the null hypothesis, AND if part of the pattern of the medians agrees with the research hypothesis, BUT part of the pattern of the medians does not, then the research hypothesis is partially supported.

    -- IF you retain the null hypothesis, OR you reject the null BUT NO PART of the pattern of the medians agrees with the research hypothesis, then the research hypothesis is not at all supported.

By the way: To properly determine if the hypothesized pattern of differences was found, one should perform pairwise comparisons (using the Median test); the report of the results given below makes use of these follow-up tests (although the computations are not shown).

By the way: Usually the researcher hypothesizes that there is a difference between the k conditions. Sometimes, however, the research hypothesis is that there is NO difference between the k conditions. If so, the research hypothesis and H0: are the same! When this is the case, retaining H0: provides support for the research hypothesis, whereas rejecting H0: provides evidence that research hypothesis is incorrect.

       For the example data, we would decide that the research hypothesis is partially supported, because the null hypothesis was rejected, and because, as hypothesized, the coop stores had the most fish. However, there was not a difference in the median number of fish in the private and chain stores.

Step 12 Reporting the results

       You will want to compute medians and IQR values to help describe the data before reporting the results of the significance test. With multiple-group designs it is often easier to present these data in a table. As for the other statistical tests, the report includes the "wordy" part and the statistical values upon which you based your statistical decision. If you reject H0:, be sure to describe how the groups differed, rather than just reporting that there was "a difference".

       The number of fish displayed at each type of store is summarized in Table 1. The median number of fish displayed was significantly among the three types of stores, using the Median test X² = 6.14, p<.05). Pairwise comparisons using the Median test (P = .05) revealed that, as hypothesized, Coop stores displayed the most fish. However, contrary to the hypothesis, there was no difference between the number of fish displayed by Chain and Private pet stores.

Table 1
Quartile values for the number of fish displayed by each of the three types of stores

	Store Type		
Number of Fish Displayed	Chain Store	Priately Owned	Coop Store
Mdn	20.00	17.00	35.00
Q1	12.00	17.00	31.25
Q3	21.50	22.50	40.25

C^2 Critical values of Chi-Square

df	$\alpha = .05$	$\alpha = .01$
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22
13	22.36	27.69
14	23.68	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.81
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98
25	37.65	44.31
26	38.89	45.64
27	40.11	46.96
28	41.34	48.28
29	42.56	49.59
∞	43.77	50.89