MEDIAN TEST

Median test is a special case of Pearson's chi-squared test. It is a nonparametric test that tests the null hypothesis that the medians of the populations from which two or more samples are drawn are identical. The data in each sample are assigned to two groups, one consisting of data whose values are higher than the median value in the two groups combined, and the other consisting of data whose values are at the median or below. A Pearson's chi-squared test is then used to determine whether the observed frequencies in each sample differ from expected frequencies derived from a distribution combining the two groups.

Hypothesis: Null hypothesis: $H_0: M_1 = M_2$

Alternative hypothesis: $H_1: M_1 \neq M_2$

Let α be the level of significance.

Procedure:

1. Compute the combined median of the two samples. Discard the values equal to median.
2. Determine for each group, the number of observations falling above and below the combined median and arrange the frequencies in 2x2 table as follows:

	Sample I	**Sample II**	**Total**
No. of scores above median	a	b	a+b
No. of scores below median	c	d	c+d
Total	a+c	b+d	n=a+b+c+d

3. Compute $\chi^2 = \frac{n(ad-bc)^2}{(a+c)(b+d)(a+b)(c+d)}$
4. If $\chi^2 < \chi^2_{tab}$ accept H_0 where $\chi^2_{tab} = \chi^2_{0.05,1}$

Solved problems:

Example 8:

Do urban and rural junior high school students differ with respect to their level of mental help:

Urban: 35 26 27 21 27 38 23 25 25 27 45 46 33 26 46 41

Rural: 29 50 43 22 42 47 42 32 50 37 37 34 31

Solution

H_O: $M_U = M_R$

H_1: $M_U \neq M_R$

LOS: $\alpha = 0.05$

The combined median is $\frac{33+34}{2} = 33.5$

Now we will form the frequency table as follows:

	Sample I	Sample II	Total
No. of scores above median	6	8	14
No. of scores below median	10	4	14
Total	16	12	28

$$\chi^2 = \frac{18(24-80)^2}{16\times12\times14\times14} = 2.33$$

${\chi^2}_{tab} = {\chi^2}_{0.05,1}$ =3.841. Here $\chi^2 < {\chi^2}_{tab}$ accept H_0.

MANN WHITNEY TEST

A popular nonparametric test to compare outcomes between two independent groups is the Mann Whitney U test. The Mann Whitney U test, sometimes called the Mann Whitney Wilcoxon Test or the Wilcoxon Rank Sum Test, is used to test whether two samples are likely to derive from the same population (i.e., that the two populations have the same shape). Some investigators interpret this test as comparing the medians between the two populations. Recall that the parametric test compares the means (H_0: $\mu_1=\mu_2$) between independent groups.

In contrast, the null and two-sided research hypotheses for the *nonparametric test* are stated as follows:

H_0: The two populations are equal versus (H_0: $\mu_1=\mu_2$)

H_1: The two populations are not equal. (H_0: $\mu_1\neq\mu_2$)

This test is often performed as a two-sided test and, thus, the research hypothesis indicates that the populations are not equal as opposed to specifying directionality. A one-sided research hypothesis is used if interest lies in detecting a positive or negative shift in one population as compared to the other. The procedure for the test involves pooling the observations from the two samples into one combined sample, keeping track of which sample each observation comes from, and then ranking lowest to highest from 1 to n_1+n_2, respectively.

Procedure:

1. Set up hypotheses and determine level of significance.
   H_0: The two populations are equal versus
   H_1: The two populations are not equal. α =0.05
2. Select the appropriate test statistic.
   Because APGAR scores are not normally distributed and the samples are small (n_1=8 and n_2=7), we use the Mann Whitney U test. The test statistic is U, the smaller of

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \text{and} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 and R_2 are the sums of the ranks in groups 1 and 2, respectively.

3. Set up decision rule.
   The appropriate critical value can be found in the table above. To determine the appropriate critical value we need sample sizes (n_1 and n_2) and our two-sided level of significance (α=0.05). The critical value for this test with $n_{1,}$ n_2 and α =0.05 is U_0 and the decision rule is as follows: Reject H_0 if $U \le U_0$.
4. Compute the test statistic.
   The first step is to assign ranks of 1 through 15 to the smallest through largest values in the total sample, as follows:
5. Conclusion: We reject H_0 because $U \le U_0$(Critical value for $n_{1,}$ n_2 and α).

Solved problems:

Example 9:

A new approach to prenatal care is proposed for pregnant women living in a rural community. The new program involves in-home visits during the course of pregnancy in addition to the usual or regularly scheduled visits. A pilot randomized trial with 15 pregnant women is designed to evaluate whether women who participate in the program deliver healthier babies than women receiving usual care. The outcome is the APGAR score measured 5 minutes after birth. Recall that APGAR scores range from 0 to 10 with scores of 7 or higher considered normal (healthy), 4-6 low and 0-3 critically low. The data are shown below.

Usual Care	8	7	6	2	5	8	7	3
New Program	9	9	7	8	10	9	6	

Is there statistical evidence of a difference in APGAR scores in women receiving the new and enhanced versus usual prenatal care? We run the test using the five-step approach.

Step 1: Set up hypotheses and determine level of significance.

H_0: The two populations are equal versus

H_1: The two populations are not equal. $\alpha = 0.05$

Step 2: Select the appropriate test statistic.

Because APGAR scores are not normally distributed and the samples are small ($n_1=8$ and $n_2=7$), we use the Mann Whitney U test. The test statistic is U, the smaller of

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \text{and} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 and R_2 are the sums of the ranks in groups 1 and 2, respectively.

Step 3: Set up decision rule.

The appropriate critical value can be found in the table above. To determine the appropriate critical value we need sample sizes ($n_1=8$ and $n_2=7$) and our two-sided level of significance ($\alpha=0.05$). The critical value for this test with $n_1=8$, $n_2=7$ and $\alpha = 0.05$ is 10 and the decision rule is as follows: Reject H_0 if $U \leq 10$.

Step 4: Compute the test statistic.

The first step is to assign ranks of 1 through 15 to the smallest through largest values in the total sample, as follows:

		Total Sample (Ordered Smallest to Largest)		Ranks	
Usual Care	New Program	Usual Care	New Program	Usual Care	New Program
8	9	2		1	
7	8	3		2	
6	7	5		3	
2	8	6	6	4.5	4.5
5	10	7	7	7	7
8	9	7		7	
7	6	8	8	10.5	10.5
3		8	8	10.5	10.5
			9		13.5
			9		13.5
			10		15
				R_1=45.5	R_2=74.5

Next, we sum the ranks in each group. In the usual care group, the sum of the ranks is R_1=45.5 and in the new program group, the sum of the ranks is R_2=74.5. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have

n(n+1)/2 = 15(16)/2=120 which is equal to 45.5+74.5 = 120.

We now compute U_1 and U_2, as follows:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 8(7) + \frac{8(9)}{2} - 45.5 = 46.5$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 8(7) + \frac{7(8)}{2} - 74.5 = 9.5$$

Thus, the test statistic is U=9.5.

Step 5: Conclusion: We reject H_0 because $9.5 \leq 10$. We have statistically significant evidence at α =0.05 to show that the populations of APGAR scores are not equal in women receiving usual prenatal care as compared to the new program of prenatal care.

Example 10

A clinical trial is run to assess the effectiveness of a new anti-retroviral therapy for patients with HIV. Patients are randomized to receive a standard anti-retroviral therapy (usual care) or the new anti-retroviral therapy and are monitored for 3 months. The primary outcome is viral load which represents the number of HIV copies per millilitre of blood. A total of 30 participants are randomized and the data are shown below.

Standard	7500	8000	2000	550	1250	1000	2250	6800
Therapy	3400	6300	9100	970	1040	670	400	
New	400	250	800	1400	8000	7400	1020	6000
Therapy	920	1420	2700	4200	5200	4100	undetectable	

Is there statistical evidence of a difference in viral load in patients receiving the standard versus the new anti-retroviral therapy?

Step 1: Set up hypotheses and determine level of significance.

H_0: The two populations are equal versus

H_1: The two populations are not equal. α=0.05

Step 2: Select the appropriate test statistic.

Because viral load measures are not normally distributed (with outliers as well as limits of detection (e.g., "undetectable")), we use the Mann-Whitney U test. The test statistic is U, the smaller of U_1, U_2

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \quad \text{and} \quad U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 and R_2 are the sums of the ranks in groups 1 and 2, respectively.

Step 3: Set up the decision rule.

The critical value can be found in the table of critical values based on sample sizes (n_1=n_2=15) and a two-sided level of significance (α=0.05). The critical value 64 and the decision rule is as follows: Reject H_0 if $U \leq 64$.

Step 4: Compute the test statistic.

The first step is to assign ranks of 1 through 30 to the smallest through largest values in the total sample. Note in the table below, that the "undetectable" measurement is listed first in the ordered values (smallest) and assigned a rank of 1.

		Total Sample (Ordered Smallest to Largest)		Ranks	
Standard **Anti-retroviral**	**New** **Anti-retroviral**	**Standard** **Anti-retroviral**	**New** **Anti-retroviral**	**Standard** **Anti-retroviral**	**New** **Anti-retroviral**
7500	400		undetectable		1
8000	250		250		2
2000	800	400	400	3.5	3.5
550	1400	550		5	
1250	8000	670		6	
1000	7400		800		7
2250	1020		920		8
6800	6000	970		9	
3400	920	1000		10	
6300	1420		1020		11
9100	2700	1040		12	
970	4200	1250		13	
1040	5200		1400		14
670	4100		1420		15
400	undetectable	2000		16	
		2250		17	
			2700		18
		3400		19	
			4100		20
			4200		21
			5200		22
			6000		23
		6300		24	
		6800		25	
			7400		26
		7500		27	
		8000	8000	28.5	28.5
		9100		30	
				$R_1 = 245$	$R_2 = 220$

Next, we sum the ranks in each group. In the standard anti-retroviral therapy group, the sum of the ranks is R_1=245; in the new anti-retroviral therapy group, the sum of the ranks is R_2=220. Recall that the sum of the ranks will always equal n(n+1)/2. As a check on our assignment of ranks, we have n(n+1)/2 = 30(31)/2=465 which is equal to 245+220 = 465. We now compute U_1 and U_2, as follows,

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 15(15) + \frac{15(16)}{2} - 245 = 100$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 = 15(15) + \frac{15(16)}{2} - 220 = 125$$

Thus, the test statistic is U=100.

Step 5: Conclusion.

We do not reject H_0 because 100 > 64. We do not have sufficient evidence to conclude that the treatment groups differ in viral load.

REFERENCES

1. Wikipedia

- https://en.wikipedia.org/wiki/Median_test
- https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test

2. The Wilcoxon-Mann-Whitney Test – An Introduction to Nonparametrics wilcox test – Kindle Edition by Frederick Ruland