Mann-Whitney U 2 Sample Test (a.k.a. Wilcoxon Rank Sum Test)

The (Wilcoxon-) Mann-Whitney (WMW) test is the non-parametric equivalent of a pooled 2-Sample *t*-test. The test assumes you have two independent samples from two populations, and that the samples have the same shapes and spreads, though they don't have to be symmetric. The WMW procedure is a statistical test of the difference between the two medians (η_1 and η_2) under the null hypothesis that they are equal.

Like the other non-parametric tests we have seen so far, the WMW test works on ranked data. The basic procedure is incredibly simple. Combine the two samples into one column, rank the data from smallest to largest (where 1 = smallest), break them down into their original samples and sum up the total rank scores (U) of each. If the null hypothesis is true then you would expect the two final *rank sums* to be about equal; the larger the difference between the two scores, the more likely that the difference is real. To test for significance we calculate an expected score:

$$E(U) = n_U (N + 1)/2 \tag{1}$$

Where $E(U)$ is the expectation of U, n_U is the sample size of the sample being tested, and N is the total sample size $N = n_1 + n_2$. It turns out that the difference between the observed and expected rank sums is best approximated through the use of a normal distribution; the area under the curve of a z-distribution. The numerator of the z score is as usual, but the denominator is more complex, but after a bunch of tedious algebra it turns out to be:

$$z = \frac{U - E(U)}{\sqrt{n_1 n_2 (N + 1)/12}} \tag{2}$$

The resulting *z score* is then looked up in a table as usual, remembering to adjust for one or two tails.

The WMW test establishes confidence intervals around the median of the differences between the two test samples called the *point estimate*. This is not so easy to do by hand as first you would need to calculate the point estimate, and then establish a confidence level as close to the 95% level as possible through non-linear interpolation…so, let MINITAB do it.

Let's work through an example.

The ethnohistoric indigenous peoples of the west Coast of North America maintained a hunting and gathering lifestyle, but one based primarily on predictable aquatic resources. As such these hunter-gatherer groups were much less mobile than most hunter-gatherer populations, setting up seasonal permanent villages, and developing a very complex hierarchical social structure. Although the groups along the west coast shared similar cultural traits, those to the north were generally more sedentary and "complex" than those to the south. Binford (2002) includes a variety of data on such groups. For this test we are interested in whether there is a significant difference between the mean annual population aggregations of groups along the Alaskan coast (n = 13) and the Californian coast (n = 12).

Therefore, let η_A = the median population aggregation of Alaskan coastal groups, and let η_C = the median population aggregation of Californian coastal groups. Formally, we wish to test the following null hypothesis at the a = 0.05 (95%) confidence level:

$$H_O : \eta_A - \eta_C = 0$$
$$H_A : \eta_A - \eta_C \neq 0$$

The data are as follows (number of individuals):

Alaska (*n* = 13)	**California (*n* = 12)**
197	50.5
162	50
57	557
108	42
53.5	23
55	26
77	45
39	96
66	113
48	30
121	33
79	45
309	

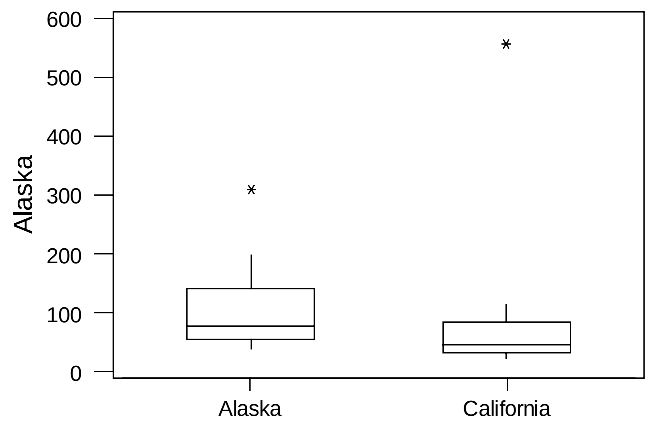
To check the assumption of similar spreads and shapes we run the descriptive statistics and produce a boxplot:

Descriptive Statistics

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
Alaska	13	105.5	77.0	93.0	77.3	21.4
California	12	92.5	45.0	53.1	148.8	43.0

Variable	Min	Max	Q1	Q3
Alaska	39.0	309.0	54.3	141.5
California	23.0	557.0	30.8	84.6

We see the mean does not equal the median, and looking at the boxplot we see both distributions are heavily skewed to the right. While the spreads are a little different, they are close enough for our purposes and we can conclude that both assumptions of the WMW are met.



Below are the calculations step by step for the hypothesis test:

1	2	3	4	5	6	7	8
Alaska	**California**	**Combined**	**Factor**	**Rank**	**Factor**	**Alaska**	**California**
197	50.5	50.5	California	11	California	23	11
162	50	50	California	10	California	22	10
57	557	557	California	25	California	14	25
108	42	42	California	6	California	19	6
53.5	23	23	California	1	California	12	1
55	26	26	California	2	California	13	2
77	45	45	California	7	California	16	7
39	96	96	California	18	California	5	18
66	113	113	California	20	California	15	20
48	30	30	California	3	California	9	3
121	33	33	California	4	California	21	4
79	45	45	California	7	California	17	7
309		197	Alaska	23	Alaska	24	
		162	Alaska	22	Alaska		
		57	Alaska	14	Alaska		
		108	Alaska	19	Alaska		
		53.5	Alaska	12	Alaska		
		55	Alaska	13	Alaska		
		77	Alaska	16	Alaska		
		39	Alaska	5	Alaska		
		66	Alaska	15	Alaska		
		48	Alaska	9	Alaska		
		121	Alaska	21	Alaska		
		79	Alaska	17	Alaska		
		309	Alaska	24	Alaska		

Columns 1 and 2: The raw data
Column 3: The two columns combined into one
Column 4: Each observation is labeled by a *factor* so that we can keep track of where it belongs
Column 5: Column 3 ranked from smallest to largest
Column 6: The factor again
Columns 7 and 8: The samples are recombined and separated into their original groups

The rank sum U_A of the Alaskan sample is 23+22+14+19+12+13+16+5+15+9+21+17+24=**210**
The rank sum U_C of the Californian sample is 11+10+25+6+1+2+7+18+20+3+4+7=**114**

As $U_A + U_C = U_T$ we can go ahead and choose one rank sum to work with, as they both will give the same result. We will use U_A. Our expected value is given by equation 1:

$$E(U) = \frac{n_U(N+1)}{2} = \frac{13(25+1)}{2} = \mathbf{169}$$

The z score is calculated using equation 2:

$$z = \frac{U - E(U)}{\sqrt{n_1 n_2 (N+1)/12}} = \frac{210-169}{\sqrt{13*12*(25+1)/12}} = \mathbf{2.23} \text{ s.d. units}$$

The 2-tailed probability associated with 2.23 s.d. units under the normal curve is $p = 0.026$. As our $p < a$ we reject the null hypothesis at the 95% level in favor of the alternative that, in fact, there is a statistically significant difference between the mean annual population aggregations of groups along the Alaskan coast and those along the Californian coast. As the rank sum for the Californian sample is much less than the Alaskan sample we could further conclude that “on average” population aggregations are larger in Alaska.

To look at the confidence limits we need to run the test through MINITAB.

>STAT
- **>NON-PARAMETRICS**
  - **>MANN-WHITNEY**
    - **>Put Alaska as the FIRST SAMPLE and California as the SECOND**
      - **>Leave the CONFIDENCE LEVEL as 95%**
        - **>Leave the ALTERNATIVE as NOT EQUAL**
          - **>OK**

The output looks as follows:

Mann-Whitney Confidence Interval and Test

```
Alaska     N =  13     Median =        77.0
California N =  12     Median =        45.0
Point estimate for ETA1-ETA2 is        27.0
95.3 Percent CI for ETA1-ETA2 is (4.5,75.0)
W = 210.0
Test of ETA1 = ETA2  vs  ETA1 not = ETA2 is significant at 0.0276
The test is significant at 0.0276 (adjusted for ties)
```

First thing to notice is that MINITAB gives us both the hypothesis test and the confidence limits without us having to run the test twice and selecting different options (I don't know why). The significance value MINITAB comes up with is $p = 0.0276$, which is slightly higher than our hand calculation ($p = 0.026$), but not enough to make any difference to the outcome. For the confidence limits we see the *point estimate* = 27, that is the estimated median of the difference between the two samples. We see MINITAB could not find us a confidence level of 95% but achieved a level of 95.3%. The lower bound is 4.5, and the upper is 75, and as they do not encompass the hypothesized value of zero, we agree with the hypothesis test and reject the null hypothesis at the $a = 0.05$ level in favor of the alternative.