

Wilcoxon-Mann-Whitney rank sum test (or U Test)

Two data samples are **independent** if they come from distinct populations and the samples do not affect each other. Using the **Mann-Whitney-Wilcoxon Test**, we can decide whether the population distributions are identical *without* assuming them to follow the normal distribution.

Example

In the data frame column **mpg** of the data set **mtcars**, there are gas mileage data of various 1974 U.S. automobiles.

```
> mtcars$mpg
 [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4
[13] 17.3 15.2 10.4 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3
[25] 19.2 27.3 26.0 30.4 15.8 19.7 15.0 21.4
```

Meanwhile, another data column in **mtcars**, named **am**, indicates the transmission type of the automobile model (0 = automatic, 1 = manual). In other words, it is the differentiating factor of the transmission type.

```
> mtcars$am
 [1] 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 1 1 1 1 1
[31] 1 1
```

In particular, the gas mileage data for manual and automatic transmissions are independent.

Problem

Without assuming the data to have normal distribution, decide at .05 significance level if the gas mileage data of manual and automatic transmissions in **mtcars** have identical data distribution.

Solution

The null hypothesis is that the gas mileage data of manual and automatic transmissions are identical populations. To test the hypothesis, we apply the **wilcox.test** function to compare the independent samples. As the p-value turns out to be 0.001817, and is less than the .05 significance level, we reject the null hypothesis.

```
> wilcox.test(mpg ~ am, data=mtcars)

wilcoxon rank sum test with continuity correction

data: mpg by am
W = 42, p-value = 0.001871
alternative hypothesis: true location shift is not equal to 0
```

Warning message:

```
In wilcox.test.default(x = c(21.4, 18.7, 18.1, 14.3, 24.4, 22.8, :
cannot compute exact p-value with ties
```

Answer

At .05 significance level, we conclude that the gas mileage data of manual and automatic transmissions in **mtcar** are *nonidentical* populations.