

Statistical Inference

Non Parametric Tests - 1

Contents

1. Non-Parametric test
2. Mann-Whitney Test
3. Wilcoxon Signed Rank test

Non-parametric statistical test

- Tests based on t and F distributions assume that populations are normally distributed.
- A large body of statistical methods is available which do not make assumptions about the nature of the distribution(e.g. normality)
- These testing procedures are termed nonparametric tests or distribution-free tests.
- If the underlying assumptions of the parametric test are met, then a parametric test will be more powerful than a non-parametric test.



Note : Always check for normality assumptions using the test explained earlier and then decide which hypothesis test is more accurate, depending upon the problem statement.

Mann-Whitney test

- The Mann-Whitney test is considered as a non-parametric alternative to t test for independent samples.
- The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed.
- The test is equivalent to Wilcoxon rank-sum test (WRS).
- The null hypothesis is that the distributions of both groups are identical, so that there is a 50% probability that an observation randomly selected from one population exceeds an observation randomly selected from another population.

Mann-Whitney test

- **Steps to follow :**

- Combine the two samples.
- Rank all the observations from smallest to largest.
- Keep track of the group to which each observation belongs.

- Tied observations (observations with same value) are assigned a rank equal to the mean of the rank positions for which they are tied.

- The test statistic is

$$U = T - \frac{m(m+1)}{2}$$

Where T is the sum of the ranks of the first sample in the combined ordered sample, m and n are sample sizes.

$$E(U) = \frac{mn}{2} \qquad V(U) = \frac{mn(m+n+1)}{12}$$

- Standardized U is assumed to follow normal distribution.
- Compare the p-value with the level of significance & conclude.

Case Study - 1

Background

Data consists of the aptitude scores of 2 groups of employees.

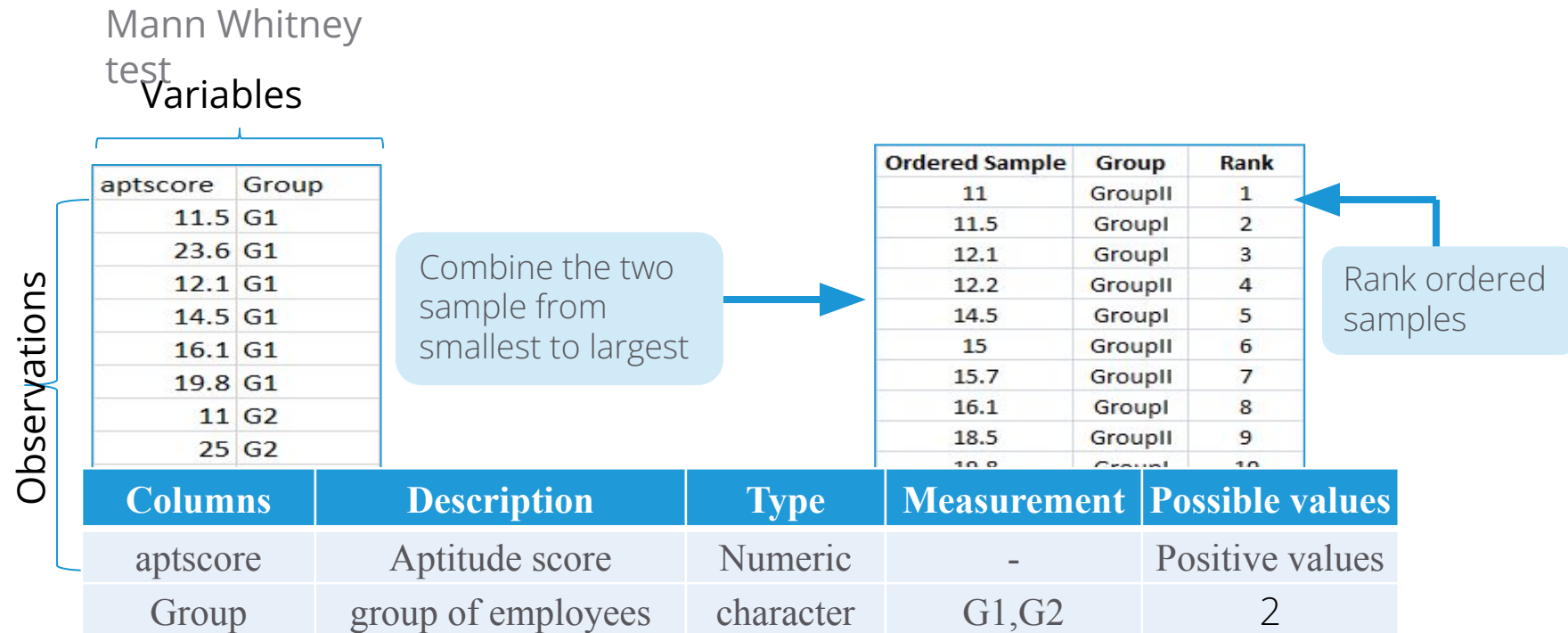
Objective

To compare aptitude scores of the two groups and test if they come from the same population.

Sample Size

Sample size: 13
Variables: aptscore, Group

Data Snapshot



- T is the sum of the ranks of the first sample in the combined ordered sample. m and n are sample sizes.

T=39 , m=6, n= 7

U=18 , E(U)=21 ,V(U)= 49

Mann-Whitney test

Testing distribution of two samples

Objective	To test the null hypothesis that the median of both samples is the same
------------------	---

Null Hypothesis (H_0): The two samples come from the same population

Alternate Hypothesis (H_1): The two samples do not come from the same population

Test Statistic	$U = T - \frac{m(m+1)}{2}$ <p>Where T is the sum of the ranks of first sample in the combined ordered sample, m and n are sample sizes</p>
Decision Criteria	Reject the null hypothesis if the p-value < 0.05

Mann-Whitney test in R

```
# Import the CSV file
```

```
data<-read.csv("Mann Whitney test.csv", header=TRUE)
```

```
# Mann-Whitney test
```

```
wilcox.test(formula=aptscore~Group,data=data)
```

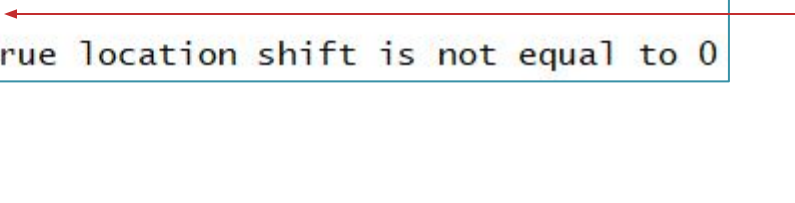


- ❑ *The Mann-whitney test is also known as the Wilcoxon Rank Sum test.*
- ❑ *The wilcox.test function gives the value of U(as W) and p-value.*
- ❑ *aptscore is the analysis variable.*
- ❑ *group is the factor.*

Mann-Whitney test in R

Output:

```
      wilcoxon rank sum test  
data:  aptscore by Group  
W = 18, p-value = 0.7308  
alternative hypothesis: true location shift is not equal to 0
```



Interpretation :

- Since p -value is >0.05 , do not reject H_0 . aptitude score is same for both the groups i.e. samples come from the same population.

Wilcoxon Signed Rank Test for paired data

- The Wilcoxon Signed Rank test is considered as a nonparametric alternative to paired t test .
- The Wilcoxon Signed Rank test is used to compare differences between two related or paired groups when the variable is either ordinal or continuous, but not normally distributed.
- H_0 : The median of difference in the population is zero
 H_1 : Not H_0 .

Wilcoxon Signed Rank Test for paired data

- **Steps to follow :**
 - Define $D_i = X_i - Y_i$, which are the differences between two values for each pair.
 - Obtain $|D_i|$, which are absolute values of differences.
 - Rank all $|D_i|$ from smallest to largest.
 - Define $R_i = \text{rank of } |D_i|$.
 - Obtain 'W', which is the sum of the ranks associated with positive D_i .
- The test statistic is W, which is the sum of the ranks associated with positive D_i . n is the sample size.

$$E(W) = \frac{n(n+1)}{4}$$

$$V(W) = \frac{n(n+1)(2n+1)}{24}$$

- Standardized W is assumed to follow normal distribution.
- Compare the p-value with the level of significance & conclude.

Case Study - 2

Background

A company organized a training program and the scores before and after training were recorded.

Objective

To test whether the median of paired samples is same.

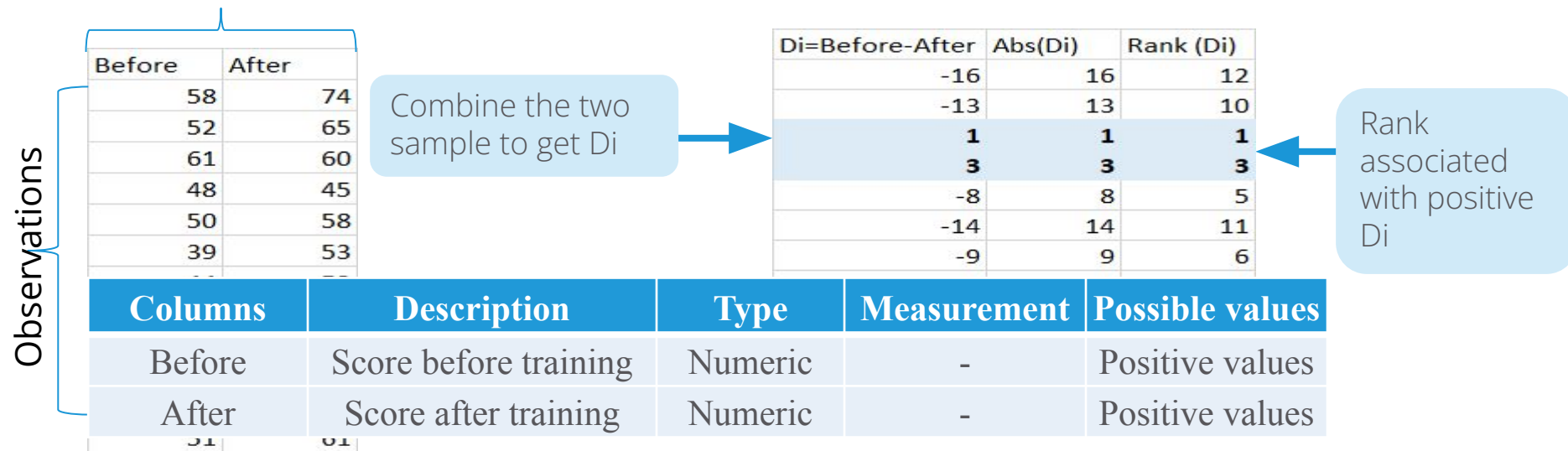
Sample Size

Sample size: 12
Variables: Before, After

Data Snapshot

- A company organized a training program and the scores before and after training were recorded.

Variables



- W is sum of the ranks associated with positive D_i . n is sample size.
 $W=4$, $n= 12$
 $E(W)=39$, $V(W)= 162.5$

Wilcoxon Signed Rank Test for paired data

Testing distribution of paired samples

Objective	To test the null hypothesis that median of paired samples is same.
------------------	--

Null Hypothesis (H_0): The median of the difference in the population is zero
Alternate Hypothesis (H_1): The median of the difference in the population is less than zero.

Test Statistic	w=sum of the ranks associated with positive D_i. $D_i = X_i - Y_i$ which are the differences between data and specified median value.
Decision Criteria	Reject the null hypothesis if p-value < 0.05

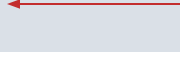
Wilcoxon Signed Rank Test for paired data in R

```
# Import the CSV file
```

```
data<-read.csv("Wilcoxon Signed Rank test for paired data.csv",  
               header=TRUE)
```

```
# Wilcoxon Signed Rank test
```

```
wilcox.test(data$Before, data$After, paired=TRUE,  
            alternative = "less")
```

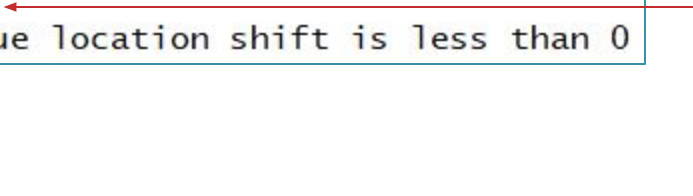


- ❑ *wilcox.test* function gives the value of W (as V) and p -value.
- ❑ *wilcox.test* function performs Wilcoxon signed rank test for paired data when `paired=TRUE` is specified.
- ❑ *Before* and *After* are the paired observations.
- ❑ `alternative=less` specifies one tail test .since, score will be more if training program is effective.

Wilcoxon Signed Rank Test for paired data in R

Output:

```
Wilcoxon signed rank test  
data: data$Before and data$After  
V = 4, p-value = 0.001709  
alternative hypothesis: true location shift is less than 0
```



Interpretation :

- Since the p -value is < 0.05 , reject H_0 . The training program is effective as the score after training is more than before training.

Quick Recap

Non Parametric Test

- Non parametric tests are performed if the normality assumption is not satisfied.

Mann-Whitney test

- Nonparametric alternative to the t test for independent samples.

Wilcoxon Signed Rank test

- Nonparametric alternative to the t test for paired samples.