

1 Nonparametric inference

In much inference and almost all that you have met so far we have assumed that the form of the distributions $f(x, \theta_1, \dots, \theta_p)$ is known except for some parameters $\theta_1, \dots, \theta_p$. In some cases we may feel that actually specifying the form of $f(x, \theta_1, \dots, \theta_p)$ is unrealistic. This is often so when the sample size is small. Our aim to construct methods of inference which do not make distributional assumptions or which are relatively insensitive to these assumptions. This leads us to nonparametric inference and to robust methods as opposed to the classical parametric methods that you met in previous. Another possibility are computationally intensive methods such as the Bootstrap.

1.1 The Kolmogorov-Smirnov test

As we have seen the chi-square goodness-of-fit tests are based on frequencies of the observations in the different classes and not the actual values of the data. Therefore all the amount of information from the data is not considered. The goodness-of-fit test of Kolmogorov-Smirnov considers all information of the raw data and applies with success to continuous distributions; it is better from the chi-square goodness-of-fit test for small samples.

Suppose that we have a random sample X_1, \dots, X_n . We wish to test the null hypothesis that the random sample is drawn from a theoretical distribution against the alternative that is not drawn from this distribution. The Kolmogorov-Smirnov statistic is,

$$D_n = \max |F_n(x) - F_0(x)|,$$

where $F_n(x)$ is the empirical distribution and $F_0(x)$ is the theoretical distribution under the null hypothesis. The empirical distribution is given by,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(X_i \leq x) = \frac{\text{number of } x_i \leq x}{n}.$$

Then the critical region of significance level α , for testing,

$$H_0 : F(x) = F_0(x) \text{ against } F(x) \neq F_0(x)$$

is $D_n \geq D_{\alpha, n}$. The critical vales of significance level α can be given using the following approximation formulas,

α	0.20	0.15	0.10	0.05	0.01
$D_{\alpha, n}$	$1.07/\sqrt{n}$	$1.14/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Example 1.1. For example suppose we have the following data set of 10 observations.

31.0 31.4 33.3 33.4 33.5 33.7 34.4 34.9 36.2 37.0

Would it be reasonable to assume that the observations are drawn from $N(\mu = 32, \sigma^2 = 3.24)$?

We will use the Kolmogorov-Smirnov goodness-of-fit statistic to check if the distributions are drawn from $N(\mu = 32, \sigma^2 = 3.24)$. The critical region of significance level α is

$$D_n \geq D_{\alpha, n} \text{ where } D_n = \max |F_n(x) - F_0(x)|.$$

Given that $F_0(x)$ is the cumulative distribution function of the $N(\mu = 32, \sigma^2 = 3.24)$,

$$F_0(x) = P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{X - 32}{\sqrt{3.24}}\right) = P\left(Z \leq \frac{X - 32}{1.80}\right)$$

and using the statistical tables we have the following results,

x	31.00	31.40	33.30	33.40	33.50	33.70	34.40	34.90	36.20	37.00
$F_0(x)$	0.29	0.37	0.76	0.78	0.80	0.83	0.91	0.95	0.99	1.00

The empirical distribution is

$$F_{10}(x) = \frac{\text{number of } x_i \leq x}{10}$$

and is given in the following table,

x	31.00	31.40	33.30	33.40	33.50	33.70	34.40	34.90	36.20	37.00
$F_{10}(x)$	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00

The Kolmogorov-Smirnov goodness-of-fit statistic is,

$$D_n = \max |F_{10}(x) - F_0(x)| = 0.4642$$

while $D_{0.05, 10} = \frac{1.36}{\sqrt{10}} = 0.43$. Therefore we reject the null hypothesis that the observations are drawn from $N(\mu = 32, \sigma^2 = 3.24)$.

1.2 Permutation Tests

The basic idea is quite simple, we find a plausible statistic, say T , for the hypotheses of interest and a family of permutations of the data such that the probability of each permutation can be found under the null hypothesis. Usually for simplicity we choose permutations which are equally likely under the null hypothesis. We then find the values of T under all possible permutations and then using this empirical distribution either find a critical region and decide whether to accept H_0 , or find the p -value, and make our decision using this value.

Example 1.2. For example suppose we have the following data set of 12 observations.

2.29828	0.59061	1.39212	2.67940	0.55242	0.05688
2.42898	4.14296	0.01564	2.82009	0.79828	1.33614

We wish to test whether the mean of the population is 1, say

$$H_0 : \mu = 1 \text{ against } H_1 : \mu \neq 1.$$

Subtracting 1, the hypothesized value of the mean from each value gives a set of deviations ($d_i, i = 1, \dots, 12$)

1.29828	-0.40939	0.39212	1.67940	-0.44758	-0.94312
1.42898	3.14296	-0.98436	1.82009	-0.20172	0.33614

One choice of statistic for our test is $T = |\bar{d}|$ of the deviations above.

Now we set up the permutations of the (modified) data:

- We first take the absolute values of all the deviations; this will give a new value for T .
- Under the null hypothesis that the median of the population is 1 the modified data are symmetric about 0, i.e., each of these deviations is equally likely to be positive or negative. Therefore a rearrangement of the data is to assign each observation to be positive or negative. Totally there 2^n permutations (rearrangements):
 1. Assign one negative sign to each value in turn and obtain n values for T .
 2. Assign two negative signs to all pairs of values and obtain more values for T .
 3. Assign minus signs to triples e.t.c., until we have the T value for all 2^n values.
- If we assume, as seems reasonable under the null hypothesis that the value of T from each permutation of signs is equally likely we have a distribution of T values. We may then choose between the hypotheses using the permutation distribution of T .

Thus if we had a sample of only 4 deviations:

$$1.29828 \quad -0.40939 \quad 0.39212 \quad 1.67940$$

we have the 16 permutations and their corresponding values for T , see Table 1.

The actual sample value of T was $T_{obs} = 0.74010$ which exceeds 10 of the 16 values, or is in the top 6 of the 16. For the definition of p -value, i.e., the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone on the assumption that the null hypothesis H_0 , is true, we have that,

$$p\text{-value} = P(T \geq T_{obs} | H_0) = \frac{\text{number of permutations with } T \geq T_{obs}}{\text{total number of permutations}}.$$

In our example $p\text{-value} = 6/16 = 0.375$ which is greater than the significance level of the test α , say 20% and thus we accept H_0 . Note in passing that the parametric test for the above problem is the one-sample t -test.

	Permutation			T
1.29828	0.40939	0.39212	1.6794	0.94480
-1.29828	0.40939	0.39212	1.6794	0.29566
1.29828	-0.40939	0.39212	1.6794	0.74010
1.29828	0.40939	-0.39212	1.6794	0.74874
1.29828	0.40939	0.39212	-1.6794	0.10510
-1.29828	-0.40939	0.39212	1.6794	0.09096
-1.29828	0.40939	-0.39212	1.6794	0.09960
-1.29828	0.40939	0.39212	-1.6794	0.54404
1.29828	-0.40939	-0.39212	1.6794	0.54404
1.29828	-0.40939	0.39212	-1.6794	0.09960
1.29828	0.40939	-0.39212	-1.6794	0.09096
-1.2982	-0.40939	-0.39212	1.679	0.10510
-1.2982	-0.40939	0.39212	-1.6794	0.74874
-1.29828	0.40939	-0.39212	-1.6794	0.74010
1.29828	-0.40939	-0.39212	-1.6794	0.29566
-1.29828	-0.40939	-0.39212	-1.6794	0.94480

Table 1: All possible permutations.

Example 1.3. 5 patients submitted to two different treatments A and B. Then the medics rate their health status based on a scale. The observations (rates) are:

Treatment A (2 patients): 1, 2

Treatment B (3 patients): 3, 5, 9

Is there any difference between the two treatments?

With other words we need to test if the two mean values are different or not,

$$H_0 : \mu_A = \mu_B \text{ against } H_1 : \mu_A \neq \mu_B.$$

Under the null hypothesis the two treatments are the same and there is no importance which treatment submitted to each patient. Therefore if we rearrange randomly the patients to the two treatments the results will be quite the same. Furthermore all rearrangements are equally likely under the null hypothesis. One choice of statistic for our test is $T = |\bar{a} - \bar{b}|$, where \bar{a} , \bar{b} are the mean values of the patients for the two treatments. If the null hypothesis is true the statistic is close to 0, while under the alternative hypothesis the statistic takes values away from 0.

We now look to rearrangements of the data observed. One possible rearrangement is 1,3 in the first sample and 2,5,9 in the second. For each rearrangement, we compute the value of T . Note that there are

$$\binom{5}{3} = 10$$

such rearrangements, see Table 2. Under the null hypothesis (that all observations come for the same distribution) all 35 rearrangements are equally likely, each with probability $1/10$.

treatment A	treatment B	\bar{a}	\bar{b}	T
1,3	2,5,9	2	5.33	3.33
1,5	2,3,9	3	4.66	1.66
1,9	2,3,5	5	3.33	1.66
1,2	3,5,9	1.5	5.66	4.51
2,3	1,5,9	2.5	5	2.5
2,5	1,3,9	3.5	4.33	0.83
2,9	1,3,5	5.5	3	2.5
3,5	1,2,9	4	4	0
3,9	1,2,5	6	2.66	3.34
5,9	1,,2,3	7	2	5

Table 2: All possible rearrangements of the patients in the two treatment groups.

The actual sample value of T was $T_{obs} = 4.51$ which exceeds 8 of the 10 values, or is in the top 2 of the 10. For the definition of p -value, i.e., the probability of getting a value of the test statistic as extreme as or more extreme than that observed by chance alone on the assumption that the null hypothesis H_0 , is true, we have that,

$$p\text{-value} = P(T \geq T_{obs} | H_0) = \frac{\text{number of permutations with } T \geq T_{obs}}{\text{total number of permutations}}.$$

In our example $p\text{-value} = 2/10 = 0.2$ which is greater than the significance level of the test α , say 10% and thus we accept H_0 . Note in passing that the parametric test for the above problem is the two-sample t -test.

Some interesting remarks are the following:

1. On the definition of p -value we have used the probability $P(T \geq T_{obs} | H_0)$. This generally speaking is not correct such also the values of the statistic which lead to rejection are small, i.e., we reject when $T \leq T_{obs}$ (two-tailed test). We have avoided this cases by selecting appropriately the test statistic using the absolute value and we reject for large values of the statistic.
2. These permutation computations are only practical for small data sets. The obvious drawback to this test is the amount of computation required. For the one sample problem with n observations we require 2^n permutations, while for the two sample model with m and n observations in the samples, there are

$$\binom{m+n}{m} = \binom{m+n}{n}$$

possible rearrangements, each with its own value of the statistic.

3. We can however come up with a very much simpler procedure at the cost of some precision. A recent suggestion is that we don't look at all permutations, but rather look a randomly chosen subset of them.

1.3 The sign test

The only assumption of the test is that the probability of one observation to be greater than δ_0 is equal to the probability that this observation is smaller than δ_0 . To this end,

$$P(X > \delta_0) = P(X < \delta_0) = \frac{1}{2}.$$

With other words δ_0 is the median of the population.

Suppose we wish to test,

$$H_0 : \delta = \delta_0$$

against

$$H_1 : \delta > \delta_0, \quad \delta < \delta_0, \quad \delta \neq \delta_0.$$

We transform the data as follows, if a value exceeds δ_0 we call it a plus while if it lies below δ_0 we call it a minus. If a value is equal to δ_0 then disregard this observation. Under H_0 the probability that a value lies above the median δ_0 and hence maps to a plus is 0.5 and the corresponding probability that we have a minus is also 0.5. Therefore the set of signs consists to a sample of binomial distribution. Finally our hypotheses are equivalent to,

$$H_0 : p = \frac{1}{2}$$

against

$$H_1 : p > \frac{1}{2}, \quad p < \frac{1}{2}, \quad p \neq \frac{1}{2}.$$

The procedure described above is known as the *sign test*. We test $H_0 : \delta = \delta_0$ by counting the number of observations that exceed the median (number of plus signs) and the number that lies below the median (number of of minus signs). Then under H_0 the number of plus signs out of the total number of signs is binomial.

The test is based on the number of positive or negative signs since we work with deviations from the median $x_i - \delta_0$. Under H_0 the probability of a plus or the proportion of positive signs is 0.5 and so the significance levels (*p*-values) can be calculated as indicated in the following steps:

1. Take the plus sign and suppose there are k of these.

2. For $H_1 : p > \frac{1}{2}$:

$$p\text{-value} = P(X \geq k | H_0 \text{ is true}),$$

for $H_1 : p < \frac{1}{2}$:

$$p\text{-value} = P(X \leq k | H_0 \text{ is true}),$$

for $H_1 : p \neq \frac{1}{2}$: If $k > n/2$,

$$p\text{-value} = P(X \geq k | H_0 \text{ is true}) + P(X \leq n - k | H_0 \text{ is true}) =$$

$$2 \times P(X \geq k | H_0 \text{ is true}) = 2 \times P(X \leq n - k | H_0 \text{ is true}),$$

else

$$p\text{-value} = 2 \times P(X \geq n - k | H_0 \text{ is true}) = 2 \times P(X \leq k | H_0 \text{ is true}).$$

The probabilities are calculated using the binomial distribution $B(n, p_0 = \frac{1}{2})$.

3. If the p -value is smaller than the significance level of the test α reject H_0 , otherwise accept H_0 .

Example 1.4. Suppose we have the following data set of 12 observations which comes from a non-normal distribution:

2.29	0.59	1.39	2.67	0.55	0.05
2.42	4.14	0.01	2.82	0.79	1.33

We wish to test whether the median of the population is 1, say

$$H_0 : \delta = 1 \text{ against } H_1 : \delta > 1.$$

We transform the data as follows, if a value exceeds 1 we call it a plus while if it lies below 1 we call it a minus. Of our data we have 7 plus signs and 5 minus signs.

The p -value is

$$P(X \geq k | p = \frac{1}{2}) = P(X \geq 7 | p = \frac{1}{2}) = 0.387$$

and so we conclude that there is no reason to reject the null hypothesis that the median is 1.

Note: For the two-tailed test ($H_1 : p \neq \frac{1}{2}$), $p\text{-value} = 2 \times 0.387$.

Suppose that the sample size n is large (greater than 25). Then we use the normal approximation to the Binomial as follows: If the proportion of plus signs is p then the test statistics

$$z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - n/2}{\sqrt{n/4}} = \frac{X - 1/2}{1/\sqrt{4n}}$$

is standard normal under the null hypothesis.

We may then base a test of $H_0 : p = 0.5$ on z . Given α we have $z_{1-\alpha}$ and $z_{1-\alpha/2}$ from normal tables and

- For $H_1 : \delta < \delta_0$ or $H_1 : p < 0.5$ reject H_0 if $z \leq -z_{1-\alpha}$
- For $H_1 : \delta > \delta_0$ or $H_1 : p > 0.5$ reject H_0 if $z \geq z_{1-\alpha}$
- For $H_1 : \delta \neq \delta_0$ or $H_1 : p \neq 0.5$ reject H_0 if $|z| \geq z_{1-\alpha/2}$

Example 1.5. Suppose we have the following data set of 30 observations which comes from a non-normal distribution:

19.5	19.8	18.9	20.4	20.2	21.5	19.9	20	18.1	20.5
18.3	19.5	18.3	19	21.7	19.5	18.2	23.9	17.5	19.7
20.9	25.3	18.4	19	23.2	17.8	24.2	19.4	20.3	23.2

Test with significance level $\alpha = 0.05$ the following hypotheses,

$$H_0 : \delta = 20.8 \text{ against } H_1 : \delta > 20.8.$$

We transform the data as follows, if a value exceeds 20.8 we call it a plus while if it lies below 20.8 we call it a minus. Of our original deviations we have 9 plus signs and 21 minus signs. Our hypotheses are equivalent to,

$$H_0 : p = \frac{1}{2} \text{ against } H_1 : p > \frac{1}{2}.$$

The test statistic is $Z = \frac{X-1/2}{1/\sqrt{(4n)}} \sim N(0, 1)$ under H_0 . We find that $z = -2.19$ and we accept H_0 at 5% since z does not lie in the critical region $z \geq z_{0.95} = 1.64$.

You will notice that the sign test makes few assumptions about the form of the parent distribution and in consequence it is a valuable tool when we have non normal distributions. However it is probably pretty clear that we must pay a price for the versatility of the test. Indeed as we replace numerical values with indicators, the signs, it seems inevitable that there will be some loss of sensitivity.

1.4 Wilcoxon's Signed Rank Test

Suppose we wish to test the same set of hypotheses,

$$H_0 : \delta = \delta_0$$

against

$$H_1 : \delta > \delta_0, \quad \delta < \delta_0, \quad \delta \neq \delta_0.$$

In the 1940's Frank Wilcoxon came up with a modification of the sign test. The test is now known as Wilcoxon's Signed Rank Test. Wilcoxon suggested that rather than map the deviations between the hypothetical median and the observations onto just two values some assessment should be made of the magnitude of these deviations. He used the rank of the deviations.

Wilcoxon took the view that a way to improve the test was to take into account the size of deviation from the median. The procedure he adopted was to take the deviations and give them a rank order ignoring the sign. Thus the smallest deviation gets rank 1 the next smallest rank 2 and so on. Deviations equally to zero are eliminated.

Now if the suggested median is the true one then the deviations should be scattered around zero and the sum of the positive ranks R_+ should be much the same as the sum of the negative ranks R_- . To test the hypothesis Wilcoxon actually worked out the probability distribution of $W = \min(R_+, R_-)$ on the assumption:

H_0 : the deviations have a symmetrical parent distribution with median zero and hence calculated the percentage points, a table of which is available.

The procedure is thus

1. Ignoring the sign assign to each difference its rank. If two or more values have the same rank then the average rank is assigned to each.
2. Given the ranks in (1) form the signed ranks by giving a negative sign to those ranks which come from negative values.
3. Compute
 R^+ = sum of the positive ranks R^- = sum of the negative ranks
4. (a) For a one tailed test, where the alternative hypothesis is that the median is more than a given value, the test statistic W is R^- .
 (b) For a one tailed test, where the alternative hypothesis is that the median is less than a given value, the test statistic W is R^+ .
 (c) For a two tailed test the test statistic W is the minimum of R^+ and R^- .
5. Reject H_0 if $W \leq W_0$, where W_0 is the critical value given in Table 12 of the Statistical Tables.

Now check that $R^+ + R^-$ are the same as $n(n+1)/2$, where n is the number in the sample (having ignored the zeros). Given the tables the test is simple and the only assumption being that the parent distribution is symmetrical. If this is not true then the sign test is your best bet.

Example 1.6. Suppose we have the following data set of 10 observations

131 127 118 135 117 112 132 120 137 113

We suspect a median of 121 so to test

$$H_0 : \delta = 121 \text{ against } H_1 : \delta \neq 121.$$

We first get the deviations from the median (under H_0)

10 6 -3 14 -4 -9 11 -1 16 -8

with 5 positive and 5 negative signs. The sign test would accept H_0 . The data and the ranks e.t.c., are laid out in Table 3.

For our data $R^+ = 38$ and $R^- = 17$. Therefore $W = \min(R^+, R^-) = 17$. Table 12 of the Statistical Tables, indicates that H_0 would be accepted for $\alpha = 0.05$ ($W_0 = 9$).

For large sample sizes, n , we use a normal approximation under the null hypothesis that the deviations D_i are independent and symmetrically distributed about zero.

Then we have that

$$E[R^+] = \frac{1}{4}n(n+1)$$

$$var(R^+) = \frac{1}{24}n(n+1)(2n+1)$$

Data	Deviation	Deviation	Rank	Signed rank
131	10	10	7	7
127	6	6	4	4
118	-3	3	2	-2
135	14	14	9	9
117	-4	4	3	-3
112	-9	9	6-	6
132	11	11	8	8
120	-1	1	1	-1
137	16	16	10	10
113	-8	8	5	-5

Table 3: The data and the ranks of Example 1.6.

and R^+ (or R^-) is asymptotically normal.

Then the z statistic,

$$z = \frac{R^+ - E(R^+)}{\sqrt{\text{var}(R^+)}} = \frac{R^+ - n(n+1)/4}{\frac{1}{24}n(n+1)(2n+1)}$$

can be used as a test statistic and we reject H_0 when $z > z_{1-\alpha}$, $z < -z_{1-\alpha}$, $|z| > z_{1-\alpha/2}$, respectively.

1.5 Paired Tests

The most common situation where the Wilcoxon or sign tests are used is when we have paired samples. Examples might be measurements made before and after treatment on the same item or paint lifetime on pairs of door on the same care. Pairing is always a good idea as it increases efficiency dramatically.

Example 1.7. Suppose we have the following data set of 10 observations.

No.	Before	After	Difference (d)
1	251	261	10
2	247	292	45
3	308	317	9
4	258	253	-5
5	267	271	4
6	256	305	49
7	230	238	8
8	268	320	52
9	269	267	-2
10	275	281	6

We could use the sign test or better still Wilcoxon's test on the differences. Our null

$$H_0 : \text{differences have zero median}$$

and the alternative

$$H_1 : \text{differences have non-zero median}$$

The procedure is, take the two samples and compute the differences. Then H_0 : no difference in medians, is the same as assuming that the differences have zero median and we can do a Wilcoxon test on the differences, giving

No.	Before	After	Difference (d)	Rank	Signed Rank
1	251	261	10	7	7
2	247	292	45	8	8
3	308	317	9	6	6
4	258	253	-5	3	-3
5	267	271	4	2	2
6	256	305	49	9	9
7	230	238	8	5	5
8	268	320	52	10	10
9	269	267	-2	1	-1
10	275	281	6	4	4

Now $R^+ = 51$ while $R^- = 4$. Therefore $W = \min(R^+, R^-) = 4$. Table 12 of the Statistical Tables, indicates that H_0 would be rejected for $\alpha = 0.05$ ($W_0 = 9$).

We could have done a sign test since as is clear we have 8+ signs in 10. The p-value is

$$2 \times P(X \geq 8) = 2 \times P(X \leq 2) = 2 \times 0.0547 = 0.1092.$$

This is on the acceptance of H_0 , illustrating the loss of efficiency which occurs with the sign test. We usually prefer the greater efficiency of Wilcoxon's test compared with the sign test assuming a symmetric parent.

1.6 Unpaired Samples - The Mann-Whitney Test

Clearly not all comparisons between samples are made with the assumption of pairing. The example below is a simple illustration.

Example 1.8. Speaking ability for patients in a study of Parkinson's disease

Had operation	2.6	2	1.7	2.7	2.5	2.6	2.5	3
Had not	1.2	1.8	1.8	2.3	1.3	3	2.2	1.3
	1.5	1.6	1.3	1.5	2.7	2		

Suppose we have two samples one from population A of size n and one from population B of size m ($m < n$) and we wish to test:

H_0 : they come from populations with the same median

against

H_1 : they come from populations with different medians

we use Mann-Whitney test as described in the steps below:

1. Rank the all data; for equal values of the observations we assign average ranks just as for Wilcoxon's signed rank test.
2. Find the sum of the ranks for the smaller sample (with m observations) say R . If $m = n$ then pick the sample you prefer.
3. Find $R' = m(m + n + 1) - R$.
4. Compute $W = \min(R, R')$.
5. Reject H_0 if W is less than the tabulated value in Table 13 or 14 of the Statistical Tables. The tables give the 5% and 2.5% points of the distribution of W . That is if the observed value of W (rounded) is less than or equal to the tabulated value then we reject H_0 the hypothesis of equal medians.

Example 1.9. Suppose we have two samples one from A of size $m = 10$ and one from B of size $n = 9$,

A	7.05	14.25	8.57	10.5	11.9	4.5	37.6	9.4	8.1	45.2
B	9.25	2.05	2.75	2.5	6.4	6.9	10	8	33	

Assign to each observation its rank in the combined set of observations as below

A	7.05	14.25	8.57	10.5	11.9	4.5	37.6	9.4	8.1	45.2
rank	7	16	10	14	15	4	18	12	9	19
B	9.25	2.05	2.75	2.5	6.4	6.9	10	8	33	
rank	11	1	3	2	5	6	13	8	17	

For the Mann-Whitney test we find the sum of the ranks for each sample $R_A = 124$ while $R_B = 66$. So, the smaller sample gives a rank sum of $R = 66$ giving $R' = 9(9 + 10 + 1) - 66 = 114$. This gives $W = \min(66, 114) = 66$ which not quite significant at 5% (two sided).

For large sample values it is conventional we can get quite good approximate values for the rank sum of the sample of size m using the normal distribution with mean and variance given by the formula below for the m A's and n B's.

$$\mu = \frac{1}{2}m(m + n + 1) \text{ and } \sigma^2 = \frac{1}{12}mn(m + n + 1).$$