Mann-Whitney Test for Independent Samples

The **Mann-Whitney U test** is essentially an alternative form of the Wilcoxon Rank-Sum test for independent samples and is completely equivalent.

Define the following test statistics for samples 1 and 2 where n_1 is the size of sample 1 and n_2 is the size of sample 2, and R_1 is the adjusted rank sum for sample 1 and R_2 is the adjusted rank sum of sample 2. It doesn't matter which sample is bigger.

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

$$U = \min(U_1, U_2)$$

As for the Wilcoxon version of the test, if the observed value of U is < U_{crit} then the test is significant (at the α level), i.e. we reject the null hypothesis. The values of U_{crit} for α = .05 (two-tailed) are given in the Mann-Whitney Tables.

Example 1: Repeat Example 1 of the Wilcoxon Rank Sum Test using the Mann-Whitney U test.

	A	B	C	D	E	F	G	H	I
1	Mann-Whitney U Test								
2									
3	Original data			Ranks					
4								Control	Drug
5	Control	Drug		Control	Drug		count	12	11
6	11	34		4	21.5		rank sum	117.5	158.5
7	15	31		10	19.5		U	92.5	39.5
8	9	35		2	23				
9	4	29		1	17		α	0.05	
10	34	28		21.5	16		tails	2	
11	17	12		11	5.5		U	39.5	
12	18	18		12.5	12.5		U-crit	33	
13	14	30		8.5	18		sig	no	
14	12	14		5.5	8.5				
15	13	22		7	14				
16	26	10		15	3				
17	31			19.5					
18	17	23.90909		117.5	158.5				

Figure 1 – Mann-Whitney U Test

Since R_1 = 117.5 and R_2 = 158.5, we can calculate U_1 and U_2 to get U = 39.5. Next we look up in the Mann-Whitney Tables for n_1 = 12 and n_2 = 11 to get U_{crit} = 33. Since 33 < 39.5, we cannot reject the null hypothesis at α = .05 level of significance.

Property 1:

$$U_1 + U_2 = n_1 n_2$$

Property 2: For n_1 and n_2 large enough the U statistic is approximately normal $N(\mu, \sigma)$ where

$$\mu = \frac{n_1 n_2}{2} \qquad \sigma^2 = \frac{n_1 n_2(n_1 + n_2 + 1)}{12}$$

Observation: Click here for proofs of Property 1 and 2.

Property 3: Where there are a number of ties, the following revised version of the variance gives better results:

$$\sigma^2 = \left(\frac{n_1 n_2}{n^2 - n}\right)\left(\frac{n^3 - n}{12} - \sum_t \frac{f_t^3 - f_t}{12}\right)$$

where $n = n_1 + n_2$, t varies over the set of tied ranks and f_t is the number of times (i.e. frequency) the rank t appears. An equivalent formula is

$$\sigma^2 = \frac{n_1 n_2 (n+1)}{12}\left(1 - \sum_t \frac{f_t^3 - f_t}{n^3 - n}\right)$$

Observation: A further complication is that it is often desirable to account for the fact that we are approximating a discrete distribution via a continuous one by applying a **continuity correction**. This is done by using a z-score of

$$z = \frac{|U - \mu| - .5}{\sigma}$$

instead of the same formula without the .5 continuity correction factor.

Example 2: Repeat Example 2 of the [Wilcoxon Rank Sum Test]() using the Mann-Whitney U test.

We show the results of the one-tailed test (without using a ties correction) is shown in Figure 2. Column W displays the formulas used in column T.

	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
3	Ranked Data									Test via normal distribution				
4														
5	Non-smokers				Smokers						Non-S	Smokers		
6	58.5	9	71	54.5	37	61.5	49.5	17.5		count	40	38		=COUNT(A6:D15)
7	15	61.5	66.5	37	56.5	2	64	54.5		median	76.5	70.5		=MEDIAN(A6:D15)
8	68.5	68.5	42	75	11	52.5	49.5	20.5		rank sum	1854	1227		=SUM(J6:M15)
9	29	44	66.5	49.5	20.5	37	15	15		U	486	1034		=T6*U6+T6*(T6+1)/2-T8
10	64	46.5	71	3.5	24.5	44	71	46.5						
11	6.5	24.5	64	44	9	40.5	5	29		α	0.05			
12	76	24.5	77	29	20.5	33	24.5	12.5		tails	1			
13	78	6.5	17.5	73.5	9	52.5	73.5	12.5		U	486			=MIN(T9:U9)
14	58.5	37	37	40.5	33	56.5	1			mean	760			=T6*U6/2
15	29	49.5	60	20.5	29	3.5	33			variance	10006.67			=T14*(T6+U6+1)/6
16										std dev	100.0333			=SQRT(T15)
17										z-score	-2.73909			=(T13-T14)/T16
18										U-crit	594.9598			=T14+T16*NORMSINV(T11)-0.5
19										p-value	0.003081			=NORMSDIST(T17)
20										sig	yes			=IF(T19<T11,"yes","no")
21										r	0.310141			=ABS(T17)/SQRT(T6+U6)

Figure 2 – Mann-Whitney U test using normal approximation

As can be seen in cell T19, the p-value for the one-tail test is the same as that found in Wilcoxon Example 2 using the Wilcoxon rank-sum test. Once again we reject the null hypothesis and conclude that non-smokers live significantly longer.

Observation: The effect size for the data using the Mann-Whitney test can be calculated in the same manner as for the Wilcoxon rank-sum test, namely

$$r = \frac{z}{\sqrt{n_1 + n_2}}$$

and the result will be the same, which for Example 2 is r = .31, as shown in cell T21.

There is another measure of effect size, namely

$$p = \frac{U}{n_1 n_2}$$

This represents the probability that a score randomly generated from population A will be bigger than a score randomly generated from population B, where A and B are the populations corresponding to the two samples and A corresponds to the sample with the higher value. The higher this value is the larger the effect.

Real Statistics Excel Functions: The following functions are provided in the Real Statistics Pack:

> **MANN**(R1, R2) = *U* for the samples contained in ranges R1 and R2

> **MANN**(R1, *n*) = *U* for the sample contained in the first *n* columns of range R1 and the sample consisting of the remaining columns in range R1. If the second argument is omitted it defaults to 1.

> **MWTEST**(R1, R2, *tails, ties, cont*) = p-value of the Mann-Whitney U test for the samples contained in ranges R1 and R2 using the normal approximation. *tails* = 1 or 2 (default). If *ties* = TRUE (default) the ties correction factor is applied. If *cont* = TRUE (default) a continuity correction is applied.

Any empty or non-numeric cells in R1 or R2 are ignored.

Observation: For Example 2, we can use the Real Statistics MANN function to arrive at the value for shown in Figure 16.3.2, namely MANN(J6:M15,N6:Q15) = 486, as well as the same p-value, namely MWTEST(J6:M15,N6:Q15,1,FALSE,TRUE) = 0.003081.

Observation: Note that the z-score and the effect size *r* can be calculated using the Real Statistics function MWTEST as follows:

> z-score = NORM.S.INV(MWTEST(R1, R2))

> *r* = NORM.S.INV(MWTEST(R1, R2))/SQRT(COUNT(R1)+COUNT(R2))

Observation: The results of analysis for Example 2 can be summarized as follows: The life expectancy of non-smokers (*Mdn* = 76.5) is significantly higher than that of smokers (*Mdn* = 70.5), $U = 486$, $z = -2.74$, $p = .0038 < .05$, $r = .31$, based on a one-tailed test Mann-Whitney test with continuity correction, but no correction for ties.

Of course, you can also use a two-tailed test with ties correction, as we will demonstrate shortly.

Real Statistics Function: The following function is provided in the Real Statistics Pack and returns output consisting of the *U*-stat, *z*-stat, *r* effect size and the three types of p-values (the normal approximation, exact test and simulation).

> **MW_TEST**(R1, R2, *lab, tails, ties, cont, exact, iter*): returns a column array with the output described above for the samples contained in ranges R1 and R2. *tails* = 1 or 2 (default). For the normal approximation, if *ties* = TRUE (default) the ties correction factor is applied; if *cont* = TRUE (default) a continuity correction is applied; if *exact* = TRUE (default FALSE) then the p-value of the exact test is output and if *iter* ≠ 0 then the p-value of the simulation version of the test is output where the simulation consists of *iter* samples (default 10,000). If *lab* = TRUE (default FALSE) then an extra column of labels is appended to the output.

Any empty or non-numeric cells in R1 or R2 are ignored. See Mann-Whitney Exact Test and Mann-Whitney Simulation for more information about the exact test and simulation p-values.

Figure 3 displays the output from =MW_TEST(A6:A17,B6:B17,TRUE) for Example 2.

	O	P	Q
10	=MW_TEST(A6:A17,B6:B17,TRUE)		
11			
12	U-stat	39.5	
13	z-stat	1.60217	
14	effect r	0.334075	
15	p-norm	0.109118	
16	p-exact	0.103709	
17	p-simul	0.108	

Figure 3 – Output from MW_TEST

Observation: Even if the argument is set to FALSE, the p-value of the exact test will be produced provided both samples have fewer than 800 elements and the smaller sample has at most 300 elements.

Real Statistics Data Analysis Tool: The Real Statistics Resource Pack also provides a data analysis tool which performs the Mann-Whitney test for independent samples, automatically calculating the medians, rank sums, U test statistic, z-score, p-values and effect size *r*.

For example, to perform the analysis in Example 1, press **Ctrl-m** and choose the **T Test and Non-parametric Equivalents** data analysis tool from the menu that appears (or from the **Misc** tab if using the Multipage user interface). The dialog box shown in Figure 4 now appears.

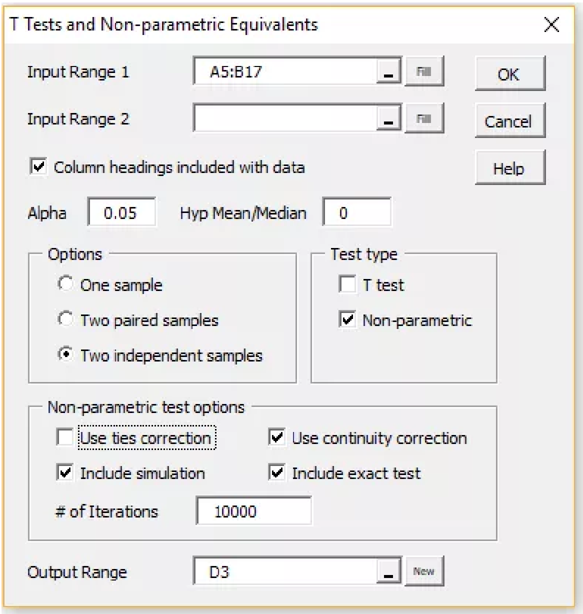

Figure 4 – Dialog box for Real Statistics Mann-Whitney Test

Enter A5:B17 as the **Input Range 1** (alternatively insert A5:A17 in **Input Range 1** and B5:B17 in **Input Range 2**), click on **Column headings included with data**, choose the **Two independent samples** and **Non-parametric** options and click on the **OK** button. Keep the default of 0 for **Hypothetical Mean/Median** and .05 for **Alpha** (although these values are not used) For this version of the test, we check **Use continuity correction**, **Include exact test** and **Include table lookup** but we leave the **Use ties correction** option unchecked.

The output is shown in Figure 5.

	A	B	C	D	E	F	G	H
3	Original data			Mann-Whitney Test for Two Independent Samples				
4								
5	Control	Drug			Control	Drug		
6	11	34		count	12	11		
7	15	31		median	14.5	28		
8	9	35		rank sum	117.5	158.5		
9	4	29		U	92.5	39.5		
10	34	28						
11	17	12			one tail	two tail		
12	18	18		U	39.5			
13	14	30		mean	66			
14	12	14		std dev	16.24808			
15	13	22		z-score	1.600189	yates		
16	26	10		effect r	0.333663			
17	31			p-norm	0.054778	0.109557		
18				p-exact	0.051854	0.103709		
19				p-simul	0.053	0.1036		

Figure 5 – Mann-Whitney test data analysis tool output

Note that both the one-tail and two-tail tests are displayed. Also, three versions of the test are shown: the test using the normal approximation (range E17:F17), the test using the exact test (range E18:F18) and the simulation test (range E19:F19). The fact that the "Yates" continuity correction factor is used is noted in cell F15.

If we check the **Use Ties correction** option in Figure 4 we would obtain the output shown in Figure 6.

	J	K	L	M	N
3	Mann-Whitney Test for Two Independent Samples				
4					
5		Control	Drug		
6	count	12	11		
7	median	14.5	28		
8	rank sum	117.5	158.5		
9	U	92.5	39.5		
10					
11		one tail	two tail		
12	U	39.5			
13	mean	66			
14	std dev	16.228	ties		
15	z-score	1.60217	yates		
16	effect r	0.334075			
17	p-norm	0.054559	0.109118		
18	p-exact	0.051854	0.103709		
19	p-simul	0.055	0.1115		

Figure 6 – Mann-Whitney test data analysis tool with ties correction

In this case the ties correction of Property 3 is applied to the normal approximation. As you can see there is very little difference between the outputs shown in Figure 5 and 6.

Note too that the ties correction (as well as the continuity correction) only applies to the normal approximation. The ties and continuity corrections are not applied to the exact and simulation versions of the test. The difference in the simulation p-values (row 19) in Figure 5 and 6 is due to the randomness of the simulations and not the ties correction.

Real Statistics Function: The Real Statistics Pack provides the following function to calculate the ties correction used in the data analysis tool.

> **TiesCorrection**(R1, R2, *type*) = ties correction value for the data in range R1 and optionally range R2, where *type* = 0: one sample, *type* = 1: paired sample and *type* = 2: independent samples

For the Mann-Whitney test *type* = 2. The ties correction is used in the calculation of the standard deviation (cell U15 of Figure 6) as follows

=SQRT(K13*(K6+L6+1)/6*(1-TiesCorrection(A6:A17,B6:B17,2)/((K6+L6)^3-K6-L6)))

Exact Test

Click here for a description of the exact version of the Mann-Whitney Test using the permutation function.

Simulation

Click here for a description of how to use simulation to determine the p-value for the Mann-Whitney test. This approach takes ties into account.

Confidence Interval of the Median

Click here for a description of how to calculate a confidence interval of the median based on the Mann-Whitney Test.

91 Responses to *Mann-Whitney Test for Independent Samples*

Oren Ben Harim *says:*
March 20, 2018 at 12:44 am

Thanks Charles, this is interesting, I started thinking of methods to improve the performance 🙂

Regard the normal approximation, I find the following:

You use the following formula:
U=min(U1,U2)
$z=(|U-\mu| -0.5)/\sigma$
By using absolute, we lose the option for distinguishing between left tale and right tale.
the test always assumes that H0 assumption is: bigger variable in the sample is bigger than the smaller variable.

I found that using $Z=(U2-\mu+c)/\sigma$ { when $U > \mu$: $c = -0.5$, when $U < \mu$: $c = 0.5$}
give you the option to choose between left and right tails.
like any other z test.

Oren

Reply

Oren *says:*
March 15, 2018 at 11:38 pm

Hi Charles

Thanks for being so helpful 🙂

1. It is probably determined by the size of the table…
2. log sound good.
3. Can you please direct me to a faster approach? is it accurate as the tables?

Thanks,
Oren

Reply

Charles *says:*
March 16, 2018 at 7:02 am

Oren,
Glad I could help. I haven't yet researched a faster approach.
Charles

Reply

Oren *says:*
March 15, 2018 at 11:58 am

Dear Charles,

I have several questions today 🙂

When is it reasonable to use the normal approximation? when n1+n1>20 or when n1>20 and n2>20?

When using the tables for small n, you get the critical values, but is there a good way to estimate the p-value?

When you calculated the exact accumulate probability in "Mann Exact" sheet, you calculated U for every combination, n!/(n1!*n2!) combinations.
Is there a faster way? or must you calculate U sperately every combination?

Thanks a lot,

Oren

Reply

Charles *says:*
March 15, 2018 at 4:51 pm

Oren,
1. I have seen advice that says n1 > 20 or n2 > 20. I have seen n1 > 20 and n2 > 20. I have seen requirements that are higher, some that are lower. It is not clear.
2. The Real Statistics software estimates the p-values from the critical table using the MPROB function. You can estimate a p-value manually by various means. The simplest is to look a the table of critical values and find the value closest to your U value (for given values of n1 and n2) and then seeing what the alpha value is. The tricky part is how to interpolate (linear, log, harmonic). Generally log interpolation will yield the best result between alpha values.
3. There are faster approaches. I have not yet implemented these.
Charles

Reply

Oren *says:*
July 30, 2018 at 2:54 pm

Hi Charles,

You used a tie correction for the normal approximation.
I can see R uses ties correction also for the Mann Whitney exact test.

Do you have any idea what is ties correction for the MW exact test?

Ps a better than log interpolation for p-value, but more memory consuming, is using a full table for all the values, not only for the specific p-values.

Thanks,

Oren

Reply

Charles *says:*

July 30, 2018 at 4:06 pm

Oren,
I have some ideas for how to deal with ties, but I have decided to use simulation (instead of an exact test) in case of ties. This has been added in the latest release of the software, which came out a couple of weeks ago). In this release, I no longer use a table of critical values. Instead I use an exact test which can handle all p values, as you suggested). I have similar capabilities for the signed ranks test.
Charles

Reply

Oren *says:*

July 31, 2018 at 5:56 am

Hi Charles,

And I thought I might be able to help you …:)
I glad to hear you went to the correct direction.

What is the maximum n you handle now in the exact?
It isn't coincident that in most tables 20 is the maximum. calculate more than 23 become very long, and each unit you add multiple the longest activity's duration by 4 (like from 20*20 => 21*21).

Intuitively I would expect the MW exact with no ties to be accurate also with ties when calculating 0.5 per each equal value from the other group.

I understand that logically you have more combinations, like if n1=1,n2=2 instead of 2 combinations (a>b, b>a) without ties, there would be 3 combinations with ties (a>b,b>a,a=b) so the calculation should be with more combinations than simple MW with no ties.

Thanks,
Oren

Charles *says:*

July 31, 2018 at 9:24 am

Oren,
I appreciate any help you can give. Several people have contributed ideas, corrections and even code.
In the latest release, the exact test works well when the size of the smaller sample is at most 300 while the larger is at most 1,000. I changed the algorithm in this release and so I am now able to generate tables of much larger size.
Charles

Oren *says:*

July 31, 2018 at 4:00 pm

Currently, I suggested on step forward, but you step 3 steps 🙂

I'm really impressed, a maximum of 300*1000 is really big, and I can see it is totally accurate!

Did you create a big table? or is it on the fly calculation?

Charles *says:*

July 31, 2018 at 10:27 pm

Oren,
It is an on the fly calculation.
Charles

Michaela *says:*
March 5, 2018 at 7:05 pm

Hi Charles,
I would like to find out whether there is a statistically significant difference between my two samples (I compare e.g. Return on Equity of subsidies recipients with non-recipients, maybe it would be even better to find out whether one sample has better results than other). The samples are not normally distributed, which should mean I should use non-parametric test, according to my findings it should be suitable to use Mann Whitney test, however I don´t know if i should look at the 1-tail or 2-tail result in Real Statistics result and basically according to which criterion do I know if it is significant? (shall I look at the P-value if it is less or more than Alpha, or shall I compare the U with the U critical?) Thank you very much in advance.
Regards, Michaela

Reply

Charles *says:*
March 7, 2018 at 9:06 pm

Michaela,
1. Whether to use a one-tailed or two-tailed test really depends on how much a priori knowledge you have about the two populations. If you know that it is impossible for the mean of population 1 to be greater than the mean of population 2, then you can use a one-tailed test. Otherwise, you should use a two-tailed test. Since it is difficult to have such a priori knowledge, usually a two-tailed test is used.
2. "shall I look at the P-value if it is less or more than Alpha, or shall I compare the U with the U critical?" These approaches are equivalent and so you can choose either one and should get the same result.
Charles

Reply

Michaela *says:*
March 10, 2018 at 11:00 am

Thank you very much Charles.

Reply

Vera *says:*
October 26, 2017 at 12:08 pm

Hi Charles,
Very interesting website. Thanks. I try to apply to my own experiments
I have three groups (independent samples (each comming from a different individual)) to compare for one variable.
sample size of each group varies from 3 to 7
I thought to use Kruskall Wallis then compare 2 groups by 2 groups using Man Whitney.
Is it correct?
or do I need to have a minimum 5 sample per groups.

Could you help me?
Thanks in advance.
best
Vera

Reply

Charles *says:*
October 27, 2017 at 10:12 am

Hi Vera,
Use of Kruskal Wallis is appropriate instead of Mann-Whitney with more than 2 variables. Of course, you could use ANOVA instead if the assumptions are satisfied. If the reason for not using ANOVA is that homogeneity of variances assumption is violated, then Welch's Test may be a better choice than KW.
Follow-up testing after KW (if you get a significant result) can be done with Mann-Whitney, but better choices are probably available. See the following webpage regarding this:
[Kruskal-Wallis Follow-up Testing]
Charles

Reply

---

difference *says:*
August 30, 2017 at 1:38 pm

Thanks for sharing your thoughts about mann-whitney.
Regards

Reply

Jake *says:*
August 10, 2017 at 9:04 am

Hi Charles, what a helpful website! Thank you!

I wonder whether I can use Mann-Whitney U test for more than 2 variables. Not only "smoker" and "non-smoker", let's say there is also "social smoker" and even more than that. I look forward to your reply. Thanks!

Reply

Charles *says:*
August 10, 2017 at 1:09 pm

Jake,
With more than two variables you would use ANOVA. The non-parametric counterpart to Mann-Whitney with more than two variables is Kruskal-wallis.
Charles

Reply

---

muhammad hashim khan *says:*
May 18, 2017 at 8:26 pm

dear charles,
i am very glad that you are performing the real life means helping the others
i am going to conduct a study on the challenges facing secondary school head teachers due to terrorism in Bannu divission KPK (both male and female schools). the sample size are 30 female and 69 are male head teachers, the data was collected on a questionaire by five likard scale SDA,DA,U,A,SAso you are requested to guide me in usind the mann whitney u test to find the difference between the male and female head teacher manualy and on spss please

Reply

Charles *says:*
May 19, 2017 at 8:26 am

Muhammad,
I don't use SPSS and so can't give you any advice about this. The referenced webpage shows how to do this manually.
Charles

Reply

---

kris *says:*

January 26, 2017 at 8:31 pm

Dear Charles,

Can you please explain the difference between the results of Mann-Whitney and Mann- Whitney with continuity correction? According to the first, a difference is not significant. But, according to the second, it is significant.

How should the results be interpreted?

Thanks,
K

Reply

Charles *says:*
January 27, 2017 at 9:46 am

Kris,
There seems to be disagreement as to whether a continuity correction should be used. Some say yes, some say no. In any case, if you have a situation where the difference is significant without a continuity correction and not significant with a continuity correction (or vice versa), then the significant result is marginal anyway and should be reported as such. There is nothing magical about alpha = .05, and a result of .049 is really not much different from a result of .051. It is not the case that the small different between these two is so important.
Charles

Reply

Paul *says:*
December 29, 2016 at 2:29 pm

Hi Charles,

The assumption of Mann Whitney test :
– Indepedent samples
– At least ordinal data
Is is correct ?

Reply

Charles *says:*
January 1, 2017 at 5:22 pm

Paul,
Yes that is correct, but for certain analyses you also have the assumption that the two samples have the same shape.
Charles

Reply

FRANCISCO *says:*
December 11, 2016 at 6:33 am

Hi,

I performed Mann whitney test in R for testing total length by sex. My sample size is 1071 (male= 523 & female=548). Im using MW due to my data dont fit a normal distribution. But, as a result I got a U = 130430. Could that be possible?

Many thanks

Reply

Charles *says:*
December 11, 2016 at 8:04 am

Francisco,
I can't say for sure, but I would doubt that R would give the wrong answer, especially for such a popular test.
Charles

Reply

Rais *says:*
October 22, 2016 at 8:13 pm

dear charles,
i performed mann whitney test on spss version 20.0. i got mann whitney test U value in thousands (e.g. U=1453) . my sample size was n=1280. i am confused to see this large value of U?? can this U value be in thousands?? U value may be in thousands?? please

Reply

Charles *says:*
October 23, 2016 at 8:12 am

Rais,
Yes, U can be this large.
Charles

Reply

Sergiy *says:*
October 12, 2016 at 2:45 pm

Dear Charles,

I have two samples which have the sample size n1=102 and n2=110. I want to compare these samples using the U-test. Could you, please, tell me how can I calculate the critical value for the Mann-Whitney U-test with given confidence level (0.05 or 0.01)?

Thanks a lot
Sergiy

Reply

Charles *says:*
October 12, 2016 at 3:38 pm

Sergiy,
This is explained on the referenced webpage, namely with such a large sample you should use the normal approximation and therefore use the critical value for the appropriate normal distribution.
Charles

Reply

Abhijit *says:*
October 3, 2016 at 5:17 pm

Hi
I performed a Wilcoxon rank-sum test with two samples x and y with sample size n_x=55 and n_y=20.
I have used matlab for this.
[p,h,stats] = ranksum(x,y)

Results are
p = 0.2678;
h = 0;
stats =

zval: 1.1082
ranksum: 853

What is the interpretation of this result? what does p value signify here?
Thanking you in advance.

Reply

Charles *says:*
October 4, 2016 at 9:38 am

Hello Abhijit,
The p value indicates that you can't reject the null hypothesis. See the following webpage for how to interpret the p-value
Null and Alternative Hypothesis.
Charles

Reply

Sergio *says:*
May 30, 2016 at 3:47 pm

Regarding the Effect Size calculation for Mann-Whitney U = (Z score / SQRT(N)), I am struggling to find any supporting reference to cite it in my thesis.

Any help please ? Thanks

Reply

Charles *says:*
May 30, 2016 at 7:39 pm

Sergio,
See http://comp.uark.edu/~whlevine/psyc5133/fritz.morris.richler.2012.xge.pdf
Charles

Reply

Sergio *says:*
May 30, 2016 at 9:18 pm

Thank you so much, Charles

Reply

Rhonda *says:*
May 17, 2016 at 4:38 pm

Hi,
First of all congratulation on your website.
I hope you can help me better understand the statistics behind the Mann Whitney test.
I have two very different sample sizes (n1 : 57, n2 : 4). Since n1+n2>20 U statistic should be considered normal with with U= n1*n2/2.
What I don't get is how the p value which is then compared to 0.05 is obtained from the U statistic. What is the relation between the U statistic, p -value and the type 1 error alpha ?
Thank you in advance

Reply

Charles *says:*
May 25, 2016 at 2:32 pm

Rhonda,
The U value is not n1*n2/2. The mean of the normal approximation is n1*n2/2 and the variance is n1*n2*(n1+n2+1)/12. You still need to calculate U as described on the referenced webpage and then p-value = 1-NORM.DIST(U, mean, stdev, TRUE).
Charles

Reply

Pingback: *Some Assembly Required » Blog Archive » Support page for GDC16 "TLDR statistics"*

Franck Limonier *says:*
March 17, 2016 at 10:18 am

Hello Charles,

Firstly, thank you for this clear and useful website !

Concerning Mann-Whitney U (M-W U) test application, a distinction is generally made between distributions of "same shape" or "different shape". In the first case, M-W U compares the medians while in the second case, it compares mean ranks.

Does Real statistics enable to test for distribution shape ? Is there a link to the 3 different versions (norm/table/exact) displayed in the data analysis tool output ?

Also, I'm not sure I get how to choose between those 3 versions, with or without ties/continuity corrections … Which procedure do you follow generally to choose the most appropriated version (in order to state only one p-value at the end) ?

thank you,

Franck

Reply

Charles *says:*
March 17, 2016 at 1:17 pm

Franck,

Your point about the shape issue is correct, but currently the Real Statistics software does not provide any test to determine whether the shape is significantly different. You can compare charts or histograms of both samples to see if they have the same shape.

The three versions of the test (norm, table, exact) are described on the website. Here is my advice as to which one to use.

For large samples, the normal approximation gives good results; if there are lots of ties then it is better to use the ties correction. There is not universal agreement as to whether or not to use the continuity correction; usually it won't make a big difference either way. I tend to always use the ties correction (since if there are no ties the results are the same). I usually don't use the continuity correction.

For small samples you should use the exact test or table of critical values. You shouldn't use the normal approximation.

If there are no or few ties and the samples are small enough so that you can use the exact test, then I would favor that over the normal approximation, although the results should be very similar. The limitation of the exact test is that it is computationally intensive and doesn't take ties into account.

Charles

Reply

Franck Limonier *says:*
March 17, 2016 at 3:44 pm

Thanks for your answer !

If distributions have different shape, are the M-W U results obtained with Real statistics still correct ? I read that SPSS used two different test procedures (depending on same vs different shape) ?

Reply

Charles *says:*
March 18, 2016 at 6:51 am

Franck,
Yes. The Real Statistics software does not take shape into account when reporting the results.
I don't know whether SPSS does or not.
Charles

Reply

Ben *says:*
February 9, 2016 at 6:19 pm

Hello Charles,

thank you very much for this excellent explanation of the U-test. It really helped me a lot to understand the entire concept and also to pull it off in Excel.

Calculating my results however, I end up with an z-value of 14,41. (I transfered the result to a z-value, since I have n= 5314).

I am quite sure I did the calulations right. Is there a chance that you have a look at the Excel and help me out?

Thank you very much

Best regards,

Ben

Reply

Charles *says:*
February 10, 2016 at 9:50 am

Ben,
Yes. Please send an Excel file with your calculations to my email address as listed on Contact Us.
Charles

Reply

Ben *says:*
February 13, 2016 at 10:34 pm

Hey Charles,

I sent you a mail. I hope it reached you.

Thanks a lot.

Ben

Reply

Charles *says:*

February 14, 2016 at 8:57 am

Ben,
I have received your email.
Charles

Reply

Ruth *says:*
November 18, 2015 at 8:43 pm

Hi Charles,

This website and tool are excellent!
I have a naive question regarding the test corrections.
I have a very small sample size: 2 groups with 3 data points each.
The two groups are clearly different between each other, but the 3 values from each group are very similar. What correction (ties, continuity) should be the most appropriate, if any?
Thank you very much.
Ruth

Reply

Charles *says:*
November 24, 2015 at 11:48 am

Ruth,
Glad you like the website and tools.
With such a small sample, it probably doesn't matter much since any result will be somewhat suspect. I would use a continuity correction and probably a ties correction as well.
Charles

Reply

Antony *says:*
September 5, 2015 at 7:05 pm

Hi,
congratulations for the website! It is really interesting.
I would ask you what is the good interpretation of the results of Wilcoxon-Mann-Whitney's test and the null hypothesis.

I have two independent samples of 15 and 10 observations respectively that describe two different types of banks. I would like to use this test to verified if belong to one of the two groups is different from belong to the other group. Do you think that this test is useful to reach this result?

Reply

Charles *says:*
September 9, 2015 at 9:14 am

The Wilcoxon-Mann-Whitney test can be used for this purpose. Generally this test is used when the assumptions for the two sample t test are not met.
Charles

Reply

Ali *says:*
July 14, 2015 at 6:13 am

Great site,
Can I use this method to compare two different tests on the same individuals (compare the results from the entire

populations of both test to see if there is a significant difference). I am currently using the receiver operating characteristics (ROC) analysis for each method to figure out which testing method can provide higher area under the curve but trying to link this to WMW test. Please help

Reply

Charles *says:*
July 14, 2015 at 7:50 pm

Ali,
The Mann-Whitney test assumes independent samples, which by definition excludes the situation where both variables apply to the same individuals. You should consider using the paired t test or the nonparametric Wilcoxon signed ranks test.
Charles

Reply

Herman *says:*
July 1, 2015 at 3:50 pm

Great resource…thanks a bunch. When I'm ready to do Mann- WHitney pairwise comparisons because the Kurskal-Wallis (non parametric ANOVA equivilent) model came back significant, must I adjust the alpha level by dividing by the number of pairwise comparisons I will be doing (as in a bonferroni adjustment) or just accept the default 0.05 alpha level?

Also, should I be reporting the the Mann-Whitney significance or that of the exact test (as I have had at least one data set come back contradictory)?

Thanks for your consideration!

Reply

Charles *says:*
July 3, 2015 at 8:56 am

Herman,

Thanks for your kind words about the Real Statistics resources.

If you are going to perform multiple post-hoc tests, you should correct for familywise error in some way (e.g. Bonferroni). In the next release of the Real Statistics Resource Pack I will add the Nemenyi post-hoc test which is like Tukey's HSD test but for Kruskal-Wallis. This test will also correct for familywise error.

If the sample size is very small (under 10) then you shouldn't use the normal approximation to the Mann-Whitney test, and so you should only report the exact version of the test. If you have lots of ties then unless your sample is very small, you shouldn't use the exact test. In all other cases, if you are getting contradictory results from the Mann-Whitney test (normal approximation vs. exact test), you should report both results. If the results are very different, then there is probably an error. Otherwise, you need to show that the test is at the borderline between significant and not significant.

Charles

Reply

Amjad *says:*
August 31, 2016 at 8:24 pm

Hi Charles,

I'm using Nemenyi test, and in some cases I get a negative p-value back. Is there a problem with the test in your package?

Thanks a lot for this wonderful resource.

Amjad

Reply

Charles *says:*
September 1, 2016 at 7:34 am

Amjad,
That is strange. Can you send me an Excel file with your data and analysis? You can find my email address at Contact us.
Charles

Reply

Car *says:*
June 26, 2015 at 2:02 am

I am trying to understand how the critical values for U are calculated. When I look up Ucrit for alpha(2)=0.05, n1=06, n2=12 in Zar 4th edition, I get 58, but your chart gives 14, and when I run the Mann Whitney U test for independent samples in Real Statistics it reports 24.072. What am I not understanding?

Thanks, Car

Reply

Charles *says:*
June 26, 2015 at 8:54 am

Car,

I believe that the values in Zar's book are equal to n1*n2-crit where crit are the values shown in the table on the Real Statistics website.

I don't know where you obtained the value 24.072. It looks like some value based on the normal approximation, but it doesn't look like it comes from some table on my site. The value of the Real Statistics formula MCRIT(6,12,.05,2) = 14.

Charles

Reply

Fiona *says:*
June 21, 2015 at 10:06 pm

Hi Charles

When I am using the tool, the rank sum which it is calculating is significantly bigger than when I ask Excel to calculate the sum of the same values in the column. Any idea why that might be occurring?

Thanks

Reply

Charles *says:*
June 22, 2015 at 8:48 am

Fiona,
If you send me an Excel file with your data, I will try to figure out what is going wrong.
See Contact Us for my email address.
Charles

Reply

Ferra *says:*
June 19, 2015 at 1:42 am

Hi,
Thank you for this useful tutorial. I have problem with using MANN_TEST. The problem is that the results of using MANN_TEST do not match the results of MANN(R1, R2)!! Even the sample counts are not correct when I use MANN_TEST…What is the problem?

Thank you again,
Ferra

Reply

Ferra *says:*
June 19, 2015 at 4:57 pm

I found the reason! I forgot to remove the check for "column headings included with data".

Thank you

Reply

Mark Stone *says:*
June 17, 2015 at 4:14 pm

Firstly thank you for this wonderful website. Things are so well explained!
Secondly, I have a question for which the answer may be so obvious that I am ashamed to ask! But in figure 2, calculate the variance you use the formula
VARIANCE=(N1*N2/2)*((N1+N2+1)/6)
I just wonder where the '6' at the end comes from?
Apologies in the advance if this is a daft question, or if it answered elsewhere (I have looked but couldn't find anything. I do tend to get lost quite easily when dealing with formulas and numbers though!

Reply

Charles *says:*
June 17, 2015 at 5:44 pm

Mark,
The answer is not so obvious. If you look at the proof of Property 2 on the webpage Mann-Whitney Test – Advanced you'll see that the 6 is the result of some mathematical calculations.
Charles

Reply

Anna *says:*
April 2, 2015 at 1:11 pm

Dear Charles,

First let me thank you for this website! It has been very useful – although I am just using the method/approach, not the computing programme.
Secondly, however, I am finding it difficult to understand my results/if i've calculated everything, or need more…

I have n1=37 and n2=37, with R1=1195.5 and R2= 1579.5…this gives me U1=876.5 and U2=492.5.
Now I know my 'expected U value' is 684.5, and I have a Stdiv of 92.5 and a Z value of -2.08

What do I take from this?

I know I can't use a critical value table as the n>20 means its normalised… but what does this mean? Is there a critical value I can work out/get a reference-able source for?

Is there a difference between my data (what shows me this?)? What actually is Z (does it matter?)...do I need to calculate P, how would I do this?

Sorry for so many questions. I know I have glaring gaps in my understanding but I just haven't been able to find anything on the web that explains things clearly (laymans English not this stats talk) for n>20 situations… Hope you can help!

Kind regards

Anna

Reply

Charles *says:*

April 3, 2015 at 8:23 pm

Anna,

For samples sufficiently large you simply use one hypothesis testing using the normal distribution as described in Example 2, with additional details shown on the webpage:

http://www.real-statistics.com/sampling-distributions/single-sample-hypothesis-testing/

Charles

Reply

Philip Campbell *says:*

March 31, 2015 at 4:18 pm

Dear Charles,

Thank you for the excellent resources pack. I am running the MANN_TEST function. I have no problem if I turn off correction but I have a lot of tied data so need it on. However, when on it returns "#VALUE!" for std dev so rest cannot be calculated. I cannot see why the correction for ties will not work. Please help.

I'm using the Excel 2007 version of the software.

Thank you,

Phil

(P.S. – I have a question regarding the Fisher Test function – is it accurate for larger then 2X2 tables – it returns a value for my 2×8 table but I'm not sure if its accurate.)

Reply

Charles *says:*

April 2, 2015 at 10:03 am

Philip,

If you send me an Excel file with your data I can try to figure out why it is returning #VALUE! for standard deviation when the Ties correction option is used.

The Fisher Exact Test as implemented only supports 2 x 2 tables.

Charles

Reply

Andrey Panchenko *says:*

March 14, 2015 at 8:37 am

Dear Charles,

I am trying to do Mann-Whitney test using data from Your example 1, but I get an error "A runtime error has occured. The analysis tool will be aborted. Type mismatch."

The results looks as follow:
count 12 11
median 14,5 28
rank sum 117,5 158,5
U 92,5 39,5

and rest are blank
one tail two tail
alpha

I have excel 2013. Could You provide any suggestion? Thank You.

Reply

Charles *says:*
March 14, 2015 at 6:46 pm

Andrey,
The usual problem is the setting of the value for alpha on the dialog box. This value defaults to .05, but for some languages you will need to re-enter the value as .05 or ,05.
Charles

Reply

Andrey Panchenko *says:*
March 14, 2015 at 7:24 pm

Charles,
Many thanks, that was the solution. Thank You for quick response and the Real-Statistics pakage.

Reply

---

nick *says:*
February 25, 2015 at 7:57 pm

I have the big R1+R2 result can I get my r1 and r2 from that?

Reply

Charles *says:*
February 25, 2015 at 11:05 pm

Nick,
I don't know what r1 and r2 are, but note that R1 + R2 = n(n+1)/2 where n = n1 + n2.
Charles

Reply

Brandon *says:*
February 16, 2015 at 7:40 am

Awesome website Charles. I'm taking reservoir characterization and your tutorials have really helped. I was trying to follow this tutorial and apply it to my data. Here's the question: Using the Mann-Whitney test, does the fracture height documented by the initial analysis of scan lines A & B represent the same population of fractures, or different populations? The scan line measurements are at different intervals for A and B. For instance scan line A has distance from origin measurements of 0, 0.6, 1.8, and 4.4. With corresponding fracture heights of 0.4, 5, 12.6, and 9.6. Scan line B has distance from origin of 0.8, 3, 4.6, and 6.4 with corresponding fracture heights of 9.6, 4.8, 11.4 and 5.5. Is their a way to use MANN_TEST for these values. I realize that I must rank my fracture heights in relation to the distance from origin. So for the values I gave it would be: 0.4, 5, 9.6, 12.6, 4.8, 9.6, 11.4, 5.5. Scan line A no has a rank sum of 13 and B has a sum of 23. What would be the best way to perform a Mann-Whitney test?

Reply

Charles *says:*
February 18, 2015 at 7:48 pm

Brandon,
If I understand your question properly, you should be able to use Mann-Whitney for this analysis using the approach described on the referenced webpage, including the MANN_TEST formula or the T Test and Non-parametric Equivalents data analysis tool.
Charles

Reply

Nina *says:*
December 17, 2014 at 1:51 pm

Never mind, it just occurred to me the p-value for a 2-tailed test is (probably) twice the p-value of the one-tailed test 🙂

Reply

Nina *says:*
December 17, 2014 at 12:48 pm

Hi Charles,
my compliments on the blog!
I was wondering if you could perhaps explain how the formulas in Example 2 change if you would like to calculate the values for a 2-tailed test.
Best
Nina

Reply

Simao C Rodeia *says:*
December 13, 2014 at 2:40 pm

Hi,

first of all, congratulation on the website. It's really well built, as least for me, a null at stats. And it's surely an enormous amount of work that you've put at our disposition.
Thank you!

Unfortunately, i'm running an older version of excel, who does not support your package.
So, trying do go trough this test with samples n1=31 and n2=15, i'm having difficulty finding the critical value. furthermore, my results do not comply with the U1+U2 = n1n2 property… actually, U1 = U2 = n1n2…
i've gone through every formula and i did not find any mistake…

Could you help me?
Thanks in advance. Cheers,
Simao

Reply

Charles *says:*
December 14, 2014 at 10:25 pm

Simao,
1. It is very strange that your results don't comply with U1+U2 = n1*n2, since mathematically this should always hold. If you send me a spreadsheet with your calculations, I will try to figure out where the problem is.
2. The critical value for alpha = .05, two-tailed, n1 = 31 and n3 = 15 is 148.
3. The Real Statistics Resource Pack works with all Windows versions of Excel back to Excel 2002. Do you have an older version than this or are you using a version for the Mac prior to Excel 2011?
Charles

Reply

student *says:*
November 3, 2014 at 5:39 am

sir, what stat tool i'm going to use if i have two groups with unequal number of respondents? I want to determine if there is a significant difference in their performance in terms of the knowing, applying and reasoning skills of the students between the control and experimental groups?

Reply

student *says:*
November 3, 2014 at 5:40 am

with less than 40 respondents each group. thanks

Reply

student *says:*
November 3, 2014 at 5:44 am

Mann-Whitney U-test?

Reply

Charles *says:*
November 4, 2014 at 8:14 am

Provided the assumptions (normality, etc.) hold then you can use the t test for independent samples. If these assumptions are violated then you can use the Mann-Whitney U test.

Reply

Felix *says:*
September 2, 2014 at 4:53 pm

Hi and thank you for all your work! Your website is an amazing resource for me.

As Fig. 4 shows there are two different output values for significance, rows 20 and 23. Where's the difference? I got different results for some of my analyses and don't know how to deal with it…

Felix

Reply

Charles *says:*
September 3, 2014 at 7:52 am

Hi Felix,

Row 20 is based on the normal approximation (when the sample size is large), while row 23 is based on the exact value using the table of critical values. If the sample is large (sample size > 20) then no row 23 is generated. If row 23 is generated then you should use the results from row 23; otherwise you should use the results from row 20 (the only choice).

Charles

Reply

Saad *says:*
July 16, 2014 at 5:15 am

Hi Charles

I noticed that calculating p-value using
MTEST(R1, R2, t) or
Ctrl-m and choose the T Test and Non-parametric Equivalents

gives different results (~ 10% different)!
Any insight pls?
Thanks

Reply

Charles *says:*
July 17, 2014 at 11:06 pm

Saad,
I have never seen this before. It sounds lie an error. Can you send me an example where this is the case?
Charles

Reply

Charles *says:*
July 20, 2014 at 10:08 pm

Hi Saad,
In al the examples that I have seen, the function and data analysis tool give the same results. Can you send me the example where the two results are different?
Charles

Reply

Amber Davidson *says:*
April 4, 2014 at 11:26 pm

Any chance you can get the mean and standard deviation for two tailed Mann-Whitney U Test? I assume when it says Wilcoxon Signed-Rank Test for Paired Samples after I do the test it is actually the Mann-Whitney U Test, correct?

And what if I'm dealing with time? Do I still leave the Mean/Median at 0? Just want to make sure it doesn't mess up my results.

And if this works, you are a LIFESAVER!

Reply

Charles *says:*
April 8, 2014 at 8:01 am

Amber,

The mean and standard deviation provided work for both the one-tail and two-tail tests. I just didn't write the information twice (e.g. in Figure 4) since it is the same.

The Wilcoxon Signed-Rank Test for Paired Samples is not the same as the Mann-Whitney U Test, although they have many characteristics in common. If you have paired samples you should use the test described on the webpage http://www.real-statistics.com/non-parametric-tests/wilcoxon-signed-ranks-test/.

The Hypothetical Mean/Median field is not used with the current implementation of the Mann-Whitney Test or Wilcoxon Signed-Rank Test for Paired Samples, and so you may assume that the value is 0.

Charles

Reply