Mann-Whitney tests in SPSS (Quiz)

The development of this E-Book has been supported by the British Academy.

Mann Whitney test practical questions

In this example, you will test whether boys tend to agree more strongly that they are interested in learning about science (the variable INTEREST_4), as well as engaging in more science-related activities than girls. The variable INTEREST_4 is an ordinal Likert scale – taking four categories ranging from Strongly Disagree to Strongly Agree – so a non-parametric Mann-Whitney is an appropriate choice of test here.

The first step is generally to test for the normality of the variable **INTEREST_4** in each of the two groups that are indicated by **GENDER**, so do this now. (In practice you might skip this step if your dependent variable is ordinal rather than continuous, but let's work through it here for completeness.)

- Question: Looking at the histogram and the normality tests does this variable look normal for **GENDER** = **Female**?
- Question: How about for **GENDER** = **Male**?

Next perform the Mann Whitney test in SPSS to test for differences in the distribution of **Interested in learning science[INTEREST_4]** between groups defined by **Student gender[GENDER]** and answer the following questions:

- Question: Which of the two groups has the highest mean rank?
- Question: Looking at the second SPSS table does it look like the two groups are significantly different?
- Question: How would you write up these findings?

The development of this E-Book has been supported by the British Academy.

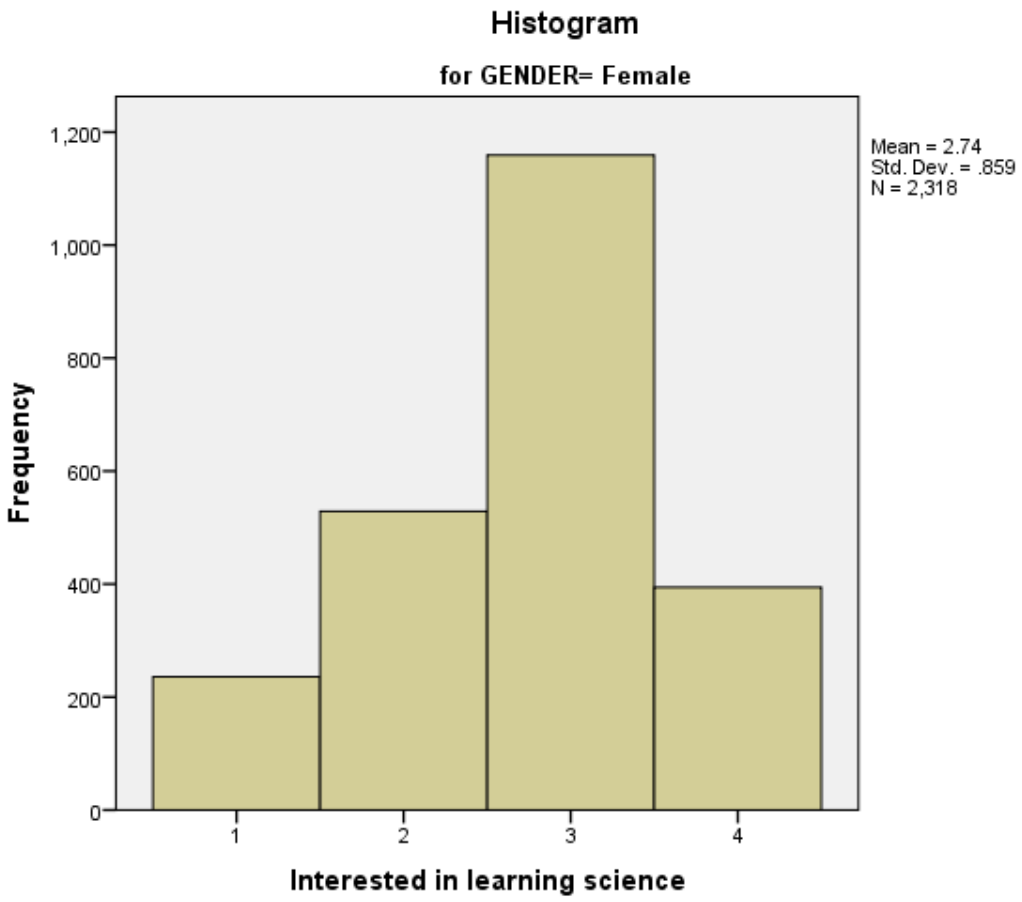
Solutions to Mann Whitney test practical questions

To test the dependent variable for normality in each group:

- Select **Descriptive Statistics** from the **Analyze** menu.
- Select **Explore** from the **Descriptive Statistics** sub-menu.
- Click on the **Reset** button.
- Copy the **Interested in learning science[INTEREST_4]** variables into the **Dependent List:** box.
- Copy the **Student gender[GENDER]** variable into the **Factor List:** box.
- Click on the **Plots...** button.
- On the screen that appears select the **Histogram** tick box.
- Unselect the **Stem and leaf** button.
- Select the **Normality plots with tests** button.
- Click on the **Continue** button.
- Click on the **OK** button.

Question: Looking at the histogram and the normality tests does this variable look normal for **GENDER** = **Female**?

Solution: The output from SPSS is as follows:



Ideally for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case 2.74. This distribution appears to be significantly skewed to the left (negatively skewed).

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Student gender	Statistic	df	Sig.	Statistic	df	Sig.
Interested in learning science	Female	.290	2318	.000	.853	2318	.000
	Male	.298	2481	.000	.841	2481	.000

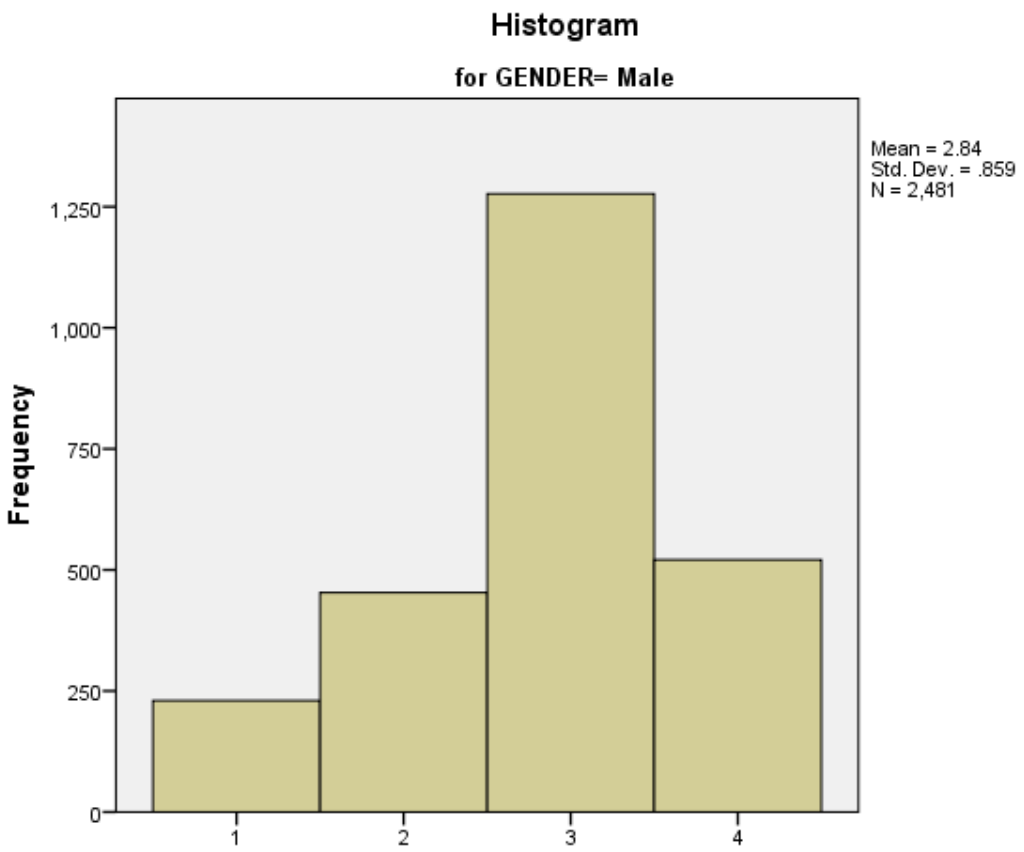
a. Lilliefors Significance Correction

For the first group: The Kolmogorov-Smirnov test has degrees of freedom which equals the number of data points, namely 2318.

Here we see that the p value is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution.

Question: How about for **GENDER** = **Male**?

Solution: The output from SPSS is as follows:



Ideally for a normal distribution this histogram should look symmetric around the mean of the distribution, in this case 2.84. This distribution appears to be significantly skewed to the left.

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Student gender	Statistic	df	Sig.	Statistic	df	Sig.
Interested in learning science	Female	.290	2318	.000	.853	2318	.000
	Male	.298	2481	.000	.841	2481	.000

a. Lilliefors Significance Correction

For the second group: The Kolmogorov-Smirnov test has degrees of freedom which equals the number of data points, namely 2481.

Here we see that the p value (quoted under Sig. for Kolmogorov Smirnov) is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the variable follows a normal distribution.
To perform the Mann Whitney test in SPSS:

- Select **Non Parametric Tests** from the **Analyze** menu.
- Select **Legacy Dialogs** from the **Non Parametric Tests** sub-menu.
- Select **2 Independent-Samples...** from the **Legacy Dialogs** sub-menu.
- Click on the **Reset** button.
- Copy the **Interested in learning science[INTEREST_4]** variable into the **Test Variable List:** box.
- Copy the **Student gender[GENDER]** variable into the **Grouping Variable:** box.
- Click on the **Define Groups...** button.
- Type **1** into the **Group 1** box.
- Type **2** into the **Group 2** box.
- Click on the **Continue** button.
- Click on the **Exact...** button.
- On the screen that appears select the **Exact** button.
- Click on the **Continue** button.
- Click on the **OK** button.

- Question: Which of the two groups has the highest mean rank?

Solution: The output from SPSS is as follows:

Ranks

	Student gender	N	Mean Rank	Sum of Ranks
Interested in learning science	Female	2318	2313.57	5362845.00
	Male	2481	2480.76	6154755.00
	Total	4799		

Here we see that for **GENDER** category Female we have 2318 observations whose total sum of ranks is 5362845.00. This results in a mean rank of 2313.57. By contrast for **GENDER** category Male we have 2481 observations whose total sum of ranks is 6154755.00. This results in a mean rank of 2480.76. So **GENDER** category Male has a larger mean rank than **GENDER** category Female and thus tends to take larger values.

- Question: Looking at the second SPSS table does it look like the two groups are significantly different?

Solution: The output from SPSS is as follows:

Test Statistics

	Interested in learning science
Mann-Whitney U	2675124.000
Wilcoxon W	5362845.000
Z	-4.524
Asymp. Sig. (2-tailed)	.000
Exact Sig. (2-tailed)	.000
Exact Sig. (1-tailed)	.000
Point Probability	.000

Here we see that the p value, quoted next to Asymp. Sig. (2-tailed) is .000 (reported as $p < .001$) which is less than 0.05. We therefore have significant evidence to reject the null hypothesis that the distribution of **INTEREST_4** is the same for both groups. In this case **GENDER** category Male has a higher mean rank than **GENDER** category Female, meaning more scores in a higher position. The asymptotic significance used above is only an approximation to the p value and it is possible to construct the exact p value. This is given in the next row and we see that the exact p value is .000 (reported as $p < .001$) whilst the asymptotic p value is also reported as .000. The exact p value agrees with the asymptotic p value that the null hypothesis can be rejected.

- Question: How would you write up these findings?

Solution:

In conclusion, we could report this to a reader as follows:

A comparison of the mean of the distribution of the variable INTEREST_4 was desired for **GENDER** categories Female and Male but due to the non-normality of the variable a Mann Whitney test was carried out. **GENDER** category Male (N= 2481) has a larger mean rank (2480.76) than **GENDER** category Female (N= 2318) with mean rank (2313.57) and thus tends to take larger values. A statistically significant difference was found ($U = 2675124.000$, $p < .001$).
As we might have predicted, boys tend to report stronger agreement with the statement that they are interested in learning about science than girls.