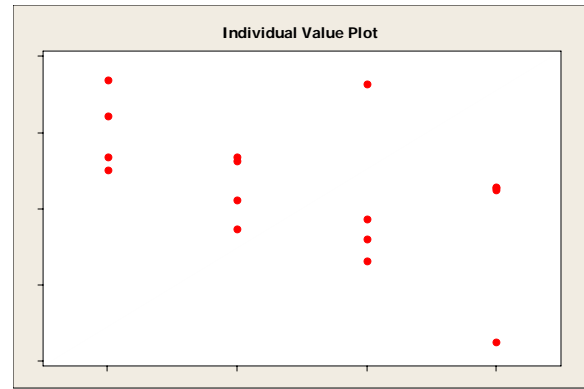


CHAPTER

26



Nonparametric Tests

Wilcoxon Rank Sum Test

Minitab does a two-sample rank test (often called the Mann-Whitney test or the two-sample Wilcoxon rank sum test) for the difference between two population medians, and calculates the corresponding point estimate and confidence interval. To perform the Wilcoxon rank sum test select

Stat > Nonparametrics > Mann-Whitney

from the menu.

Example 26.1 describes the following setting for the Wilcoxon rank sum test. A researcher planted corn in 8 plots of ground, then weeded the corn to allow no weeds in 4 plots and exactly 3 weeds per meter in the other 4 plots. The table here and in EG26-01.MTW shows yields of corn (bushels per acre) in each of the plots.

Weeds per meter	Yield (bu/acre)			
0	166.7	172.2	165.0	176.9
3	158.6	176.4	153.1	156.0

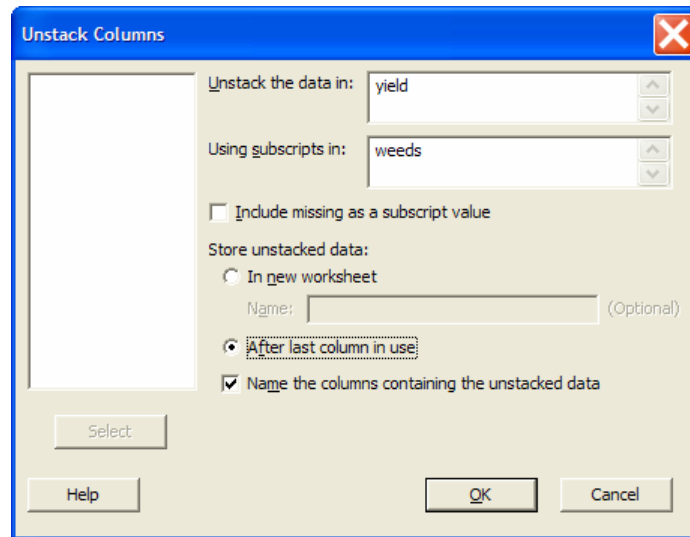
Minitab assumes that the data are independent random samples from two populations that have the same shape (hence the same variance) and a scale that is at least ordinal. The data

need not be from normal populations. The data for this command needs to be in two columns containing data from two populations. The columns do not need to be the same length.

Since the data in EG26-01.MTW are arranged with the number of weeds in one column and the yield in another, it is necessary to unstack the data in two columns. To unstack the data, select

Data ► Unstack Columns

from menu. In the dialog box, enter “Yield” after Unstack the data in and “Weeds” after Using subscripts in. Store the unstacked data after the last column in use.

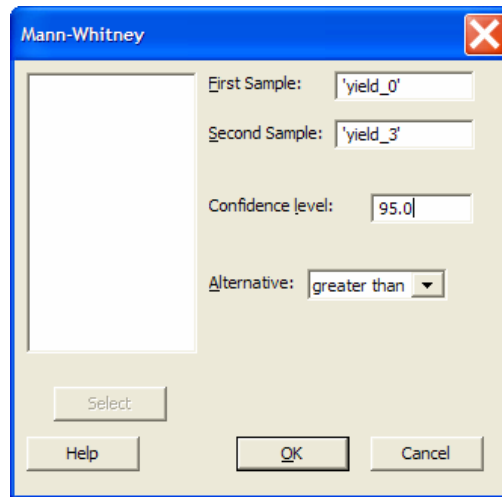


Once the new columns are added, the worksheet will contain the following data.

	C1	C2	C3	C4	C
	weeds	yield	yield_0	yield_3	
1	0	166.7	166.7	158.6	
2	0	172.2	172.2	176.4	
3	0	165.0	165.0	153.1	
4	0	176.9	176.9	156.0	
5	3	158.6			
6	3	176.4			
7	3	153.1			
8	3	156.0			

Perform the test by selecting **Stat ► Nonparametrics ► Mann-Whitney** from the menu. In the First Sample and Second Sample columns, enter the columns with sample data, select the appropriate Alternative, and click OK. The Mann-Whitney tests $H_0: \text{median}_1 = \text{median}_2$ against

a one- or two-sided alternative. Since we expect a higher yield in the weed-free plots, we select the “greater than” alternative.



Minitab determines the attained significance level of the test using a normal approximation with a continuity correction factor. We see from the output that follows that the sum of the ranks in the first group (0 weeds) is $W = 23$, with approximate P -value 0.0970. The effect of weeds on yield is not statistically significant at the 0.05 level.

Mann-Whitney Test and CI: Yield_0, Yield_3

	N	Median
Yield_0	4	169.45
Yield_3	4	157.30

Point estimate for ETA1-ETA2 is 11.30
 97.0 Percent CI for ETA1-ETA2 is (-11.40, 23.80)
 $W = 23.0$
 Test of ETA1 = ETA2 vs ETA1 > ETA2 is significant at 0.0970

Wilcoxon Signed Rank Test

Minitab performs a one-sample Wilcoxon signed rank test of the median for single samples or matched pairs. To perform the Wilcoxon signed rank test, select

Stat ► **Nonparametrics** ► **1-Sample Wilcoxon**

from the menu. The Wilcoxon test assumes that the data are a random sample from a symmetric population that is not necessarily normal.

Example 26.6 in BPS considers a study of early childhood education. Kindergarten students were asked to tell a fairy tale that had been read to them earlier in the week. Each child told two stories. The first had been read to them and the second had been read and also illustrated with pictures. An expert listened to a recording of the children and assigned a score for

certain uses of language. Here and in EG26-06.MTW are the data for five “low progress” readers in a pilot study:

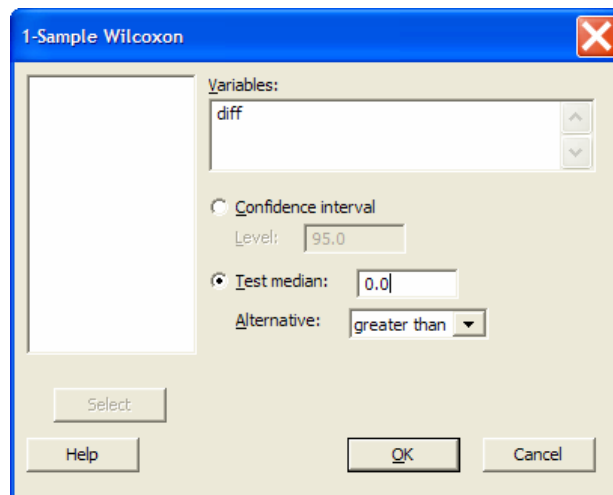
Child	1	2	3	4	5
Story 1	0.77	0.49	0.66	0.28	0.38
Story 2	0.40	0.72	0.00	0.36	0.55
Difference	0.37	-0.23	0.66	-0.08	-0.17

We will test the hypotheses

H_0 : scores have the same distribution for both stories

H_a : scores are systematically higher for Story 2

Because these are matched pairs data, we base our inference on the differences. Enter the differences into a Minitab worksheet. Select **Stat** ► **Nonparametrics** ► **1-Sample Wilcoxon** from the menu. Enter the variable and specify the null hypothesis and the alternative, as shown in the dialog box.



The output following shows that the observed value $W^+ = 9$. Minitab uses the Normal approximation with the continuity correction to give the approximate P -value of 0.394. This small sample is not statistically significant.

Wilcoxon Signed Rank Test: diff

Test of median = 0.000000 versus median > 0.000000

	N	for Test	Wilcoxon Statistic	P	Estimated Median
diff	5	5	9.0	0.394	0.1000

Kruskal-Wallis Test

The Kruskal-Wallis test is a rank test that can replace the ANOVA F test. This test is a generalization of the procedure used by Mann-Whitney. The test assumes that the data arise as k independent random samples from continuous distributions, all having the same shape (normal or not). The null hypothesis of no differences among the k populations is tested against the alternative of at least one difference. To perform the test, select

Stat ► Nonparametrics ► Kruskal-Wallis

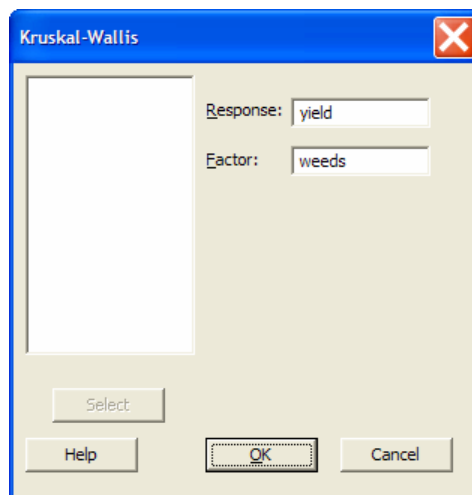
from the menu. The factor column may be numeric or text, and may contain any value. The levels do not need to be in any special order.

Example 26.10 in BPS provides an illustration of the Kruskal-Wallis test. The data in the example and EG26-10.MTW gives the number of weeks allowed to grow in each plot and the corn yield. The hypotheses tested in the example are

H_0 : yields have the same distribution in all groups

H_a : yields are systematically higher in some groups than in others

The Kruskal-Wallis test can be performed by selecting **Stat ► Nonparametrics ► Kruskal-Wallis** from the menu and filling in the columns for Response and Factor as shown.



The following output gives the results $H = 5.56$ and $P = 0.135$, which do not provide convincing evidence that weeds have an effect on yield.

Kruskal-Wallis Test: Yield versus Weeds

Kruskal-Wallis Test on Yield

Weeds	N	Median	Ave Rank	Z
0	4	169.5	13.1	2.24
1	4	163.7	8.4	-0.06
3	4	157.3	6.3	-1.09
9	4	162.6	6.3	-1.09
Overall	16		8.5	

H = 5.56 DF = 3 P = 0.135
 H = 5.57 DF = 3 P = 0.134 (adjusted for ties)

- NOTE * One or more small samples

EXERCISES

- 26.1 To detect the presence of harmful insects in farm fields, we can put up boards covered with a sticky material and examine the insects trapped on the boards. Which colors attract insects best? Experimenters placed boards of several colors at random locations in a field of oats. Here and in EX26-01.MTW are the counts of cereal leaf beetles trapped by boards colored blue and green:

Blue	16	11	20	21	14	7
Green	37	32	20	29	37	32

Because the samples are small, it is difficult to verify approximate Normality. Select **Data > Unstack Columns** to unstack the data. Then select **Stat > NonparametricsError! Bookmark not defined. > Mann-Whitney** from the menu to use the Wilcoxon rank sum test. State hypotheses, and give the P -value and your conclusions.

- 26.2 Exercise 19.13 in BPS and EX26-02.MTW report the results of a study of the effect of the pesticide DDT on nerve activity in rats. The nerve responses are

<i>DDT group</i>	12.207	16.869	25.050	22.429	8.456	20.589
Control group	11.074	9.686	12.064	9.351	8.182	6.642

- (a) Select **Graph > Individual Plots > With Groups** to make a graph of the data. What does it show?
- (b) Select **Data > Unstack Columns** to unstack the data. Then select **Stat > NonparametricsError! Bookmark not defined. > Mann-Whitney** from the menu to use the Wilcoxon rank sum test. Compare the responses in the two groups. State hypotheses and give your conclusions.
- 26.15 Table 19.1 in BPS and TA19-01.MTW give the pretest and posttest scores for two groups of students taking a program to improve their basic mathematics skills. Did the treatment group show significantly greater improvement than the control group?
- (a) Select **Calc > Calculator** to calculate the post-test versus pre-test differences. Then select **Data > Unstack Columns** to unstack the data into separate columns for the two groups of students. Finally, select **Stat > NonparametricsError!**

- Bookmark not defined. ► **Mann-Whitney** from the menu to apply the Wilcoxon rank sum test to the posttest versus pretest differences. What do you conclude?
- (b) Select **Stat** ► **Basic Statistics** ► **2-Sample t** to compare your findings with those from the two-sample t test.
- (c) What are the null and alternative hypotheses for each of the two tests we have applied to these data?
- (d) What must we assume about the data to apply each of the tests?
- 26.16 Example 26.4 in BPS describes a study of the attitudes of people attending outdoor fairs about the safety of the food served at such locations. The full data set is given in EX26-16.MTW. It contains the responses of 303 people to several questions. The variables in this data set are (in order): subject, hfair, sfair, sfast, srest, and gender. The variable sfair contains the responses described in the example concerning safety of food served at outdoor fairs and festivals. The variable srest contains responses to the same question asked about food served in restaurants. The variable gender contains 1 if the respondent is a woman, 2 if the respondent is a man. We saw that women are more concerned than men about the safety of food served at fairs. Select **Data** ► **Unstack Columns** to unstack the data into separate columns for men and women. Then select **Stat** ► **Nonparametrics** ► **Mann-Whitney** from the menu to apply the Wilcoxon rank sum test to see if this is also true for restaurants.
- 26.19 Table 18.1 in BPS and EX26.19 give matched pairs data for 21 subjects. The response variable is time to complete a maze, both wearing a scented mask and wearing an identical mask that is unscented. Does the scent improve performance (that is, shorten the time needed to complete the maze)? The matched pairs t test (Example 16.3 in BPS) works well, and gives $P = 0.365$. Select **Stat** ► **Nonparametrics** ► **1-Sample Wilcoxon** from the menu to compare the Wilcoxon signed rank test. What is the value of W^+ ? Find the P -value for the test. Does the Wilcoxon signed rank test lead to essentially the same result as the t test?
- 26.23 Exercise 18.7 of BPS and EX26-24.MTW report the following data on the percent of nitrogen in bubbles of ancient air trapped in amber:

63.4 65.0 64.4 63.3 54.8 64.5 60.8 49.1 51.0

We wonder if ancient air differs significantly from the present atmosphere, which is 78.1% nitrogen.

- (a) Graph the data, and comment on skewness and outliers. A rank test is appropriate.
- (b) We would like to test hypotheses about the median percent of nitrogen in ancient air (the population):

$$H_0: \text{median} = 78.1$$

$$H_a: \text{median} \neq 78.1$$

Select **Stat** ► **Nonparametrics** ► **1-Sample Wilcoxon** from the menu to apply the Wilcoxon signed rank test to the differences between the observations and 78.1. (This is the one-sample version of the test.) What do you conclude?

- 26.24 Table 18.3 in BPS and EX26-24.MTW give data on the healing rate (micrometers per hour) of the skin of newts under two conditions. This is a matched pairs design, with the body's natural electric field for one limb (control) and half the natural value for another limb of the same newt (experimental). We want to know if the healing rates are systematically different under the two conditions. You decide to use a rank test. Select **Stat** ► **Nonparametrics** ► **1-Sample Wilcoxon** to find the signed rank statistic W^+ and the P -value. Give a conclusion. Be sure to include a description of what the data show in addition to the test results.
- 26.25 Cola makers test new recipes for loss of sweetness during storage. Trained tasters rate the sweetness before and after storage. Here and in EX26-25.MTW are the sweetness losses (sweetness before storage minus sweetness after storage) found by 10 tasters for one new cola recipe:

2.0 0.4 0.7 2.0 -0.4 2.2 -1.3 1.2 1.1 2.3

Are these data good evidence that the cola lost sweetness?

- (a) These data are the differences from a matched pairs design. State hypotheses in terms of the median difference in the population of all tasters. Select **Stat** ► **Nonparametrics** ► **1-Sample Wilcoxon** to carry out a test. Give your conclusion.
- (b) Select **Stat** ► **Basics Statistics** ► **1-Sample t** from the menu to find the one-sample t test P -value for these data. How does this compare with your result from (a)? What are the hypotheses for the t test? What assumptions must we make for each of the t and Wilcoxon tests?
- 26.28 Example 25.6 in BPS used ANOVA to analyze the results of a study to see which of four colors best attracts cereal leaf beetles. Here and in EX26-28.MTW are the data:

Color	Insects trapped					
Yellow	45	59	48	46	38	47
White	21	12	14	17	13	17
Green	37	32	15	25	39	41
Blue	16	11	20	21	14	7

Because the samples are small, we will apply a nonparametric test. Select **Stat** ► **Nonparametrics** ► **Kruskal-Wallis** from the menu to calculate the Kruskal-Wallis statistic H . What does the test lead you to conclude?

26.29 Table 25.2 in BPS and EX26-31.MTW contains data comparing the number of trees and number of tree species in plots of land in a tropical rainforest that had never been logged with similar plots nearby that had been logged 1 year earlier and 8 years earlier. The third response variable is species richness, the number of tree species divided by the number of trees. There are low outliers in the data, and a histogram of the ANOVA residuals shows outliers as well. Because of lack of Normality in small samples, we may prefer the Kruskal-Wallis test.

- Make a graph to compare the distributions of richness for the three groups of plots. Also give the median richness for the three groups.
- Select **Stat** ► **Nonparametrics** Error! Bookmark not defined. ► **Kruskal-Wallis** from the menu to use the Kruskal-Wallis test to compare the distributions of richness. State hypotheses, the test statistic and its P -value, and your conclusions.

26.30 Here and in EX26-30.MTW are the breaking strengths (in pounds) of strips of polyester fabric buried in the ground for several lengths of time:

2 weeks	118	126	126	120	129
4 weeks	130	120	114	126	128
8 weeks	122	136	128	146	140
16 weeks	124	98	110	140	110

Breaking strength is a good measure of the extent to which the fabric has decayed. The standard deviations of the 4 samples do not meet our rule of thumb for applying ANOVA. In addition, the sample buried for 16 weeks contains an outlier. Select **Stat** ► **Nonparametrics** Error! Bookmark not defined. ► **Kruskal-Wallis** from the menu to use the nonparametric test. What are the hypotheses for the Kruskal-Wallis test, expressed in terms of medians? Report your conclusion.

26.36 Exercise 17.8 compares the number of tree species in unlogged plots in the rain forest of Borneo with the number of species in plots logged 8 years earlier. Here and in EX26-36.MTW are the data:

Unlogged	22	18	22	20	15	21	13	13	19	13	19	15
Logged	17	4	18	14	18	15	15	10	12			

- Select **Graph** ► **Stem-and-Leaf** to make stemplots of the data. Does there appear to be a difference in species counts for logged and unlogged plots?
- Does logging significantly reduce the mean number of species in a plot after 8 years? State the hypotheses. Select **Data** ► **Unstack Columns** to unstack the data. Then select **Stat** ► **Nonparametrics** Error! Bookmark not defined. ► **Mann-Whitney** from the menu to do a rank test, and state your conclusion.

26.49 A botanist prepares 16 identical planting pots and then introduces different numbers of nematodes into the pots. A tomato seedling is transplanted into each plot. Here and in

EX26-49.MTW are data on the increase in height of the seedlings (in centimeters) 16 days after planting.

Nematodes	Seedling growth			
0	10.8	9.1	13.5	9.2
1,000	11.1	11.1	8.2	11.3
5,000	5.4	4.6	7.4	5.0
10,000	5.8	5.3	3.2	7.5

- (a) What hypotheses does Kruskal-Wallis test?
- (b) Do nematodes appear to retard growth? Select **Stat** ► **Nonparametrics** ► **Kruskal-Wallis** from the menu to apply the Kruskal-Wallis test. What do you conclude?